Stochastic Processes and their Applications 10 (1980) 221-254

ASYMPTOTIC INFERENCE FOR STOCHASTIC PROCESSES

Ishwar V. BASAWA,
Department of Mathematical Statistics, La Trobe University, Australia

B.L.S. PRAKASA RAO,
Indian Statistical Institute, New Delhi, India and University of Wisconsin, Madison, WI, U.S.A.

Received 15 October, 1979
Revised 6 March, 1980

This is a survey of some aspects of large-sample inference for stochastic processes. A unified framework is used to study the asymptotic properties of tests and estimators of parameters in discrete-time, continuous-time jump-type, and diffusion processes. Two broad families of processes, viz, ergodic and non-ergodic type are introduced and the qualitative differences in the asymptotic results for the two families are discussed and illustrated with several examples. Some results on estimation and testing via Bayesian, nonparametric, and sequential methods are also surveyed briefly.

AMS (1970) subject class: 62F10, 62F20, 62M07, 62M10, 62M99
Maximum likelihood estimator; Markov processes; likelihood ratio and score tests; density estimation; ergodic and non-ergodic type processes; Bayes estimation and tests; jump type and diffusion processes; sequential methods; asymptotic efficiency of tests and estimators

1. Introduction

This paper is concerned with a survey of results on the asymptotic theory and methods of inference as applied to stochastic processes. Likelihood based methods for discrete-time, and continuous-time processes including diffusion processes will be reviewed in a unified framework. We shall also study old and new asymptotic optimality criteria for estimators and tests. Some aspects of nonparametric, sequential, and Bayesian methods for dependent observations will be briefly discussed.

Details covering most of the topics in this paper will be available in a forthcoming monograph, Basawa and Prakasa Rao (1980). The selection and coverage of topics in the present article reflect our own interest and familiarity with the work in this area. For this reason we do not claim to present a comprehensive treatment of all the main results in asymptotics for stochastic processes; on the contrary, it will be easy for the readers to notice several important omissions. We may cite two such ommissions, viz., identification and selection procedures for processes (see Bechhofer, Kiefer and Sobel (1968)), and work on simulation analysis (e.g. Crane and Lemoine (1977)) which is especially suited to stochastic processes. We have placed greater emphasis on parametric likelihood methods than on other types of approaches. An attempt is made to unify various results concerning asymptotic likelihood methods as applied to

(i) independent and identically distributed (i.i.d.) observations,
(ii) Markov processes,
(iii) arbitrary discrete-time processes and
(iv) continuous-time processes including diffusion.

Also, two broad classes of processes viz., ergodic type and non-ergodic type are discussed in a unified setup and the important qualitative differences of results for the two classes are highlighted and explained through several examples.

We begin with some historical remarks in Section 2, and Section 3 contains a brief discussion of some statistical concepts and criteria for the benefit of those who are not familiar with statistical terms. In Section 4 the results for classical i.i.d. models and their extension to Markov processes are surveyed briefly. A general model for discrete-time processes which includes both the ergodic and non-ergodic type processes is formulated in Section 5 where several examples are also discussed. Section 6 is concerned with asymptotic optimality of estimators and tests for ergodic type processes. Recent results on optimality of estimators and tests for non-ergodic type processes are summarized in Section 7. In Section 8 continuous-time processes including diffusion are discussed along with some examples. The last three sections (Sections 9 to 11) summarize some aspects of Bayesian inference, nonparametric inference and sequential methods for dependent observations.

Research in the general area of stochastic processes up to this stage is dominated by work on modelling and probabilistic analysis of the models. When one contemplates using these models to explain observed phenomena, various questions of a statistical nature arise – in particular, problems of estimation and testing hypotheses are encountered. Research on the statistical aspects of stochastic processes is of recent origin and seems to be lagging behind the theoretical probabilistic developments in the area. It is hoped that this article (and the monograph by the authors) will help in directing the interest of research workers towards the statistical methods.

2. Some remarks on historical developments

The purpose of this section is to give a broad outline of the landmarks in the developments of main ideas and methods of likelihood asymptotic inference discussed in this paper in Sections 4 to 8. Specific references on the topics covered in Sections 9 to 11 are mentioned in those respective sections. For a more detailed bibliography see the monograph, Basawa and Prakasa Rao (1980).

Many of the basic ideas and techniques of estimation originated in the fundamental papers of Fisher (1922, 1925) who was concerned with the classical model of independent and identically distributed observations. In these papers Fisher developed the *method of maximum likelihood* and introduced the important concept of *sufficiency*; large-sample properties including the *efficiency* of the maximum likelihood estimator (MLE) were also discussed in Fisher (1925). The formulation of the modern hypotheses testing problem is due to Neyman and Pearson (1928). See Lehmann (1959) for a comprehensive treatment of testing problems. A general decision-theoretic formulation of statistical problems which includes estimation and hypotheses testing was developed by Wald (1939, 1950). Cramér (1946), Wilks (1944), Rao (1961, 1962, 1963), Bahadur (1964); and Wald (a series of papers, see the collection (1958)), among others, laid a rigorous foundation to the large-sample theory of inference by providing proofs and refinements of the properties of the MLE discussed previously by Fisher and by deriving limit distributions and efficiency results concerning likelihood-ratio and related tests. The works mentioned so far were concerned with the classical model of i.i.d. observations. Billingsley (1961), among others, extended the classical large-sample theory results to Markov processes.

Le Cam (1953, 1960, also see 1974) gave a very general treatment of the large-sample theory and introduced the locally asymptotically normal (L.A.N.) family of models; the L.A.N. family in particular includes the classical i.i.d. model and ergodic type process (see Section 5).

Grenander (1950) extended the basic concepts of estimation and testing to general stochastic processes, and Bartlett (1955) discussed some applications of statistical methods to models in classical stochastic processes. Wald (1948), Bar-Shalom (1971), Prasad (1973), Prakasa Rao (1974), Bhat (1974), Crowder (1976), Basawa, Feigin and Heyde (1976), among others, established, under various regularity conditions, the large sample properties of the MLE for the ergodic type processes. Also, see M.M. Rao (1963, 1966) for a rigorous treatment of inference problems for general stochastic processes.

Weiss (1971, 1973, 1975) and Weiss and Wolfowitz (1974, and the references therein) developed some general large-sample techniques and discussed asymptotic optimality criteria for both estimation and testing. The conditions assumed by Weiss and Wolfowitz include the classical i.i.d. model and the ergodic type processes in addition to some non-standard examples. Also, the asymptotic optimality criteria developed by Weiss and Wolfowitz (1974) can be applied to non-ergodic type processes as shown by Heyde (1978), Basawa and Scott (1978) and Basawa and Koul (1979, 1980).

The work on large-sample inference for non-ergodic type processes (see Section 5) evolved through the papers of Heyde and Feigin (1975), Feigin (1975, 1976, 1978), Heyde (1978), Basawa and Scott (1977, 1979) Basawa (1977), Davies (1978), Basawa and Koul (1979, 1980).

3. Some statistical concepts and criteria

In this section we collect definitions and explanations of some common statistical terms and criteria for the benefit of those not familiar with statistics. For further details a text book such as Bickel and Doksum (1977) may be consulted.

Let $X(n)=(X_1,X_2,\ldots,X_n)$ be a vector of random variables having a (joint) density $p_n(x(n);\theta)$, $\theta\in\Theta\subset R$ (real line), with respect to a σ-finite positive product measure μ_n. $X(n)$ may be viewed as a *sample* (or realization) of n successive observations from a stochastic process $\{X_n, n\geq 1\}$. Suppose the functional form of this density $p_n(\cdot;\theta)$ is known while the value of θ is unknown. By an *inference problem* we mean a problem of drawing conclusions regarding the value of θ on the basis of the information in $X(n)$. In particular, an *estimator* $\hat{\theta}_n(X(n))$ of θ based on $X(n)$ is a measurable function $\hat{\theta}_n:\mathscr{X}(n)\to\Theta$, where $\mathscr{X}(n)$ is the *sample-space* (i.e. the set of all possible values of $X(n)$) and Θ is the *parameter-space*. Suppose we wish to test a *hypothesis* $H:\Theta\in\omega,\omega\subset\Theta$ against the *alternative hypothesis* $K:\theta\in\Theta-\omega$. Typically, a *test* (non-randomized) of H against K based on $X(n)$ is an indicator function

$$\phi_n(X(n))=\begin{cases}1, & \text{if } T_n(X(n))\in C,\ C\subset R,\\ 0, & \text{otherwise,}\end{cases}$$

where T_n is any measurable function $T_n:\mathscr{X}(n)\to R$. We identify the value '1' with the *rejection* of H and the value '0' with the *acceptance* of H. The set $\{x(n): T_n(x(n))\in C\}$ is called the *critical region* and the function T_n a *test-statistic*. $\phi_n(\cdot)$ is a *test function* or simply a *test*.

It is often required to choose a 'good' estimator (or test-statistic) from among a class of possible estimators (or tests). Several optimality criteria for estimation and tests are available in the literature. Most of these criteria are such that an 'optimum' estimator (or a test) must be a function of a *sufficient statistic* $S_n(X(n))$ when one exists. The concept of sufficiency was introduced by Fisher. The density function $p_n(x(n);\theta)$ of $X(n)$ viewed as a function in θ, say $L_n(\theta;x(n))$, is also known as a *likelihood function* of θ based on $x(n)$. A statistic S_n is said to be *sufficient* for θ if (and only if) the conditional density of $X(n)$ given S_n is free from θ. When S_n is a sufficient statistic the likelihood function $L_n(\theta;x(n))$ admits the following factorization

$$L_n(\theta;x(n))=H(x(n))G(\theta;S_n(x(n)))$$

where the function H depends only on $x(n)$ and G depends only on θ and S_n.

The theory of asymptotic inference is typically concerned with a study of the limiting properties (i.e. the properties of the limiting distributions) of estimators (or tests) as the sample-size $n\to\infty$. Several asymptotic criteria of optimality are available which help one to choose a 'good' estimator (or a test) from a class of possible estimators (tests). We define below some of these criteria which will be used in the sequel. A sequence of estimators $\{\hat{\theta}_n\}$, $n\geq 1$, of θ is said to be *consistent* for θ if $\hat{\theta}_n(X(n))\to\theta$, in probability, as $n\to\infty$. Let e be a class of consistent estimators $\hat{\theta}_n$

such that, as $n \to \infty$, for some $0 < k_n(\theta) \uparrow \infty$, $k_n(\theta)$ non-random,

$$k_n(\theta)(\hat{\theta}_n(X(n)) - \theta) \Rightarrow N(0, \sigma_{\hat{\theta}}^2(\theta)), \quad 0 < \sigma_{\hat{\theta}}^2(\theta) < \infty,$$

uniformly in θ, where $\Rightarrow$ denotes distribution convergence.

Asymptotic variance criterion. An estimator $\hat{\theta}_n^0 \in e$ is said to be asymptotically efficient (with respect to the class e) if, when $\hat{\theta}_n$ is any other estimator in e, we have $\sigma_{\hat{\theta}^0}^2(\theta) \leq \sigma_{\hat{\theta}}^2(\theta)$ for all $\theta \in \Theta$.

Weiss and Wolfowitz (1974) have developed the following more general criterion of estimation efficiency:

Let e_W denote the class of consistent estimators $\hat{\theta}_n$ such that, as $n \to \infty$

$$k_n(\theta)(\hat{\theta}_n(X(n)) - \theta) \Rightarrow L_{\hat{\theta}}(\theta)$$

uniformly in θ, where $L_{\hat{\theta}}(\theta)$ is a random variable having a distribution function $F_{\hat{\theta}}(u; \theta)$ (not necessarily normal). Assume that $F_{\hat{\theta}}$ is a continuous distribution function.

Weiss–Wolfowitz criterion (simplified version). An estimator $\hat{\theta}_n^0 \in e_W$ is said to be asymptotically efficient if when $\hat{\theta}_n$ is any other estimator in e_W, we have

$$\lim_{n\to\infty} P_{n,\theta}\{|K_n(\theta)(\hat{\theta}_n^0(X(n)) - \theta)| < a\} \geq \lim_{n\to\infty} P_{n,\theta}\{|k_n(\theta)(\hat{\theta}_n(X(n)) - \theta)| < a\}$$

for any fixed $a > 0$, and all $\theta \in \Theta$.

In general, $\hat{\theta}_n^0$ will depend on a. However, in standard applications of the type considered in this paper $\hat{\theta}_n^0$ will be free from the choice of a.

If the limiting variable $L_{\hat{\theta}}(\theta)$ has a normal distribution the class e_W for the standard applications reduces to the class e and it can be shown then that the Weiss–Wolfowitz criterion reduces to the asymptotic variance criterion. The Weiss–Wolfowitz criterion is especially useful when dealing with estimators having non-normal limiting distributions (see Section 7).

We now turn to the optimality (large-sample) criteria for tests. Let $\phi_n(T_n)$ be a test function based on a test-statistic T_n. For a suitably chosen sequence of positive constants $K_n(\theta)\uparrow\infty$ suppose $K_n(\theta)(T_n(X(n)) - \theta)$ has a limiting distribution. The *power function* of the test $\phi_n(T_n)$ is defined as

$$\beta_{T_n}(\theta) = E_\theta[\phi_n(T_n)].$$

Suppose now that we are interested in testing the hypothesis $H: \theta = \theta_0$ against $K: \theta > \theta_0$. A test $\phi_n(T_n)$ is said to be size-α if $\beta_{T_n}(\theta_0) = \alpha$, $(0 < \alpha < 1)$.

Local power criterion. A size-α test $\phi_n(T_n^0)$ is said to be asymptotically efficient if

$$\lim_{n\to\infty}\left\{K_n^{-1}(\theta_0)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\beta_{\phi_n(T_n^0)}^{(\theta)}\right)_{\theta_0}\right\} \geq \lim_{n\to\infty}\left\{K_n^{-1}(\theta_0)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\beta_{\phi_n(T_n)}^{(\theta)}\right)_{\theta_0}\right\}$$

where $\phi_n(T_n)$ is any size-α test based on a statistic T_n. (Both T_n and T_n^0 are such that $K_n(\theta)(T_n-\theta)$ and $K_n(\theta)(T_n^0-\theta)$ have limiting distributions.) See Rao (1961), and Basawa and Scott (1977) for details.

Pitman power criterion. Consider a sequence of alternative values θ_n, such that $\theta_n \to \theta_0$, as $n \to \infty$. More specifically, let $\theta_n = \theta_0 + k_n^{-1}(\theta_0)h$, where $h > 0$. A size-α test $\phi_n(T_n^0)$ is said to be asymptotically efficient (at h) if

$$\lim_{n\to\infty} \beta_{\phi_n(T_n^0)}(\theta_n) \geq \lim_{n\to\infty} \beta_{\phi_n(T_n)}(\theta_n),$$

where $\phi_n(T_n)$ is any size-α test based on a statistic T_n, and both T_n and T_n^0 are such that $k_n(\theta)(T_n-\theta)$ and $k_n(\theta)(T_n^0-\theta)$ possess limiting distributions. (See Rao (1963) and Feigin (1978) for details.)

The local and Pitman power criteria are related as follows: Let

$$A_\alpha(\theta_0) = \lim_{n\to\infty} \left\{\frac{\mathrm{d}}{\mathrm{d}\theta} \beta_{\phi_n(T_n)}(\theta)\right\}_{\theta_0} k_n^{-1}(\theta_0) \quad \text{and} \quad B_\alpha(\theta_0, h) = \lim_{n\to\infty} \beta_{\phi_n(T_n)}(\theta_n), \quad h > 0$$

be the local and the Pitman powers of a size-α test $\phi_n(T_n)$. Then, under mild regularity conditions one can show that

$$A_\alpha(\theta_0) = \lim_{h\downarrow 0}\left\{\frac{B_\alpha(\theta_0, h) - B_\alpha(\theta_0, 0)}{h}\right\}.$$

See Rao (1963) for details.

Maximum likelihood (ML) estimator. A measurable function $\hat{\theta}_n : \mathscr{X}(n) \to \Theta$ such that

$$L_n(\hat{\theta}_n; x(n)) = \sup_{\theta\in\Theta} L_n(\theta; x(n))$$

is known as a ML estimator of θ. Under regularity conditions it is often possible to compute $\hat{\theta}_n$ as an appropriate root of the *likelihood equation*

$$\mathrm{d}\log L_n(\theta; x(n))/\mathrm{d}\theta = 0.$$

It can be shown that under regularity conditions a ML estimator is asymptotically efficient both in the sense of asymptotic variance and Weiss-Wolfowitz criteria (see Sections 6 and 7).

4. Classical model and extensions to Markov processes

4.1. The classical model

First, consider the classical model of i.i.d. observations. Let $X_1, X_2, \ldots$ be i.i.d. random variables with common density $f(x; \theta)$, $\theta \in \Theta \subset R$. The likelihood function is

$$L_n(\theta; X(n)) = \prod_{i=1}^{n} f(X_i; \theta)$$

so that

$$\mathrm{d}\log L_n/\mathrm{d}\theta = \sum_{i=1}^{n} g(X_i;\theta),$$

where $g(X_i;\theta) = \mathrm{d}\log f(X_i;\theta)/\mathrm{d}\theta$, assuming differentiability of f. Much of the early work on asymptotics in this context is devoted to showing that, under regularity conditions, there exists a consistent estimator $\hat{\theta}_n(X(n))$ which satisfies the likelihood equation $\mathrm{d}\log L_n/\mathrm{d}\theta = 0$ with probability tending to 1 as $n\to\infty$ and further verifying that such a $\hat{\theta}_n$ is a local maximum of $L_n(\theta; X(n))$ with probability tending to 1 as $n\to\infty$ and is essentially unique. One then studies the limit distribution of a consistent root $\hat{\theta}_n$ of the likelihood equation. The key results needed for this are

(a) the law of large numbers and

(b) the central limit theorem for the sum $\sum_{i=1}^{n} g(X_i;\theta)$.

Consider the following regularity conditions:

(RC.1) The set of x for which $f(x;\theta)>0$ does not depend on θ. $f(x;\theta)$ is thrice differentiable with respect to θ and the derivatives f', f'', f''' are continuous in θ for any x. Furthermore, f is differentiable with respect to θ twice under the integral sign.

(RC.2) $E_\theta(g(X;\theta))^2 = i(\theta)$, say, where $0<i(\theta)<\infty$.

(RC.3) Let $G(x) = \sup_{\theta\in N} (\mathrm{d}^3/\mathrm{d}\theta^3)(\log f(x;\theta))$ where N is a neighborhood of θ_0. Then, $E_{\theta_0}\{G(X)\}<\infty$. ($\theta_0$ denotes the true value of θ.)

Under the above regularity conditions it is not difficult to show that, as $n\to\infty$,

(i) $(\mathrm{d}\log L_n/\mathrm{d}\theta)/n \to 0$, in probability;

(ii) $(\mathrm{d}\log L_n/\mathrm{d}\theta)/\sqrt{n} \Rightarrow N(0, i(\theta_0))$, where $N(a,b)$ denotes a normal random variable with mean a and variance b;

(iii) $\{(\mathrm{d}\log L_n/\mathrm{d}\theta)/\sqrt{n} - i(\theta)\sqrt{n}(\hat{\theta}_n-\theta_0)\}\to 0$,

in probability, where $\hat{\theta}_n$ is any consistent root (whose existence is assured by the above regularity conditions) of the likelihood equation.

The results (ii) and (iii) above imply that

(iv) $\sqrt{n}(\hat{\theta}_n-\theta_0)\Rightarrow \mathrm{N}(0, i^{-1}(\theta_0))$.

These are the well known classical results first stated heuristically by Fisher (1925) and later made more precise by other authors.

4.2. Markov processes

The extension of this model to Markov processes is straight-forward. Let $\{X_k, k\geq 1\}$ be a Markov process on a general state-space $\mathscr{X}$ with time-homogeneous transition measures

$$p(x, A;\theta) = P_\theta\{X_{k+1}\in A \mid X_k = x\},\quad x\in\mathscr{X},\ A\in\mathscr{F},$$

where $\mathscr{F}$ is the σ-field associated with $\mathscr{X}$. Assume that $p(x,A;\theta)$ admits a unique stationary distribution $\pi(\cdot\,;\theta)$ on $\mathscr{F}$ determined by

$$\pi(A;\theta) = \int_{\mathscr{X}} \pi(\mathrm{d}x;\theta) p(x,A;\theta)\quad \text{for all } A\in\mathscr{F}.$$

Further, suppose $p(x, A; \theta)$ admits a transition density $f(x, y; \theta)$ with respect to a measure μ. Thus,

$$p(x, A; \theta) = \int_A f(x, y; \theta)\, d\mu(y), \quad A \in \mathcal{F}.$$

The likelihood function based on the sample $(X_1, \ldots, X_n)$ is given by

$$L_n(\theta; X(n)) = f_0(x_1; \theta) \prod_{k=1}^{n-1} f(x_k, x_{k-1}; \theta),$$

where $f_0(\cdot\,; \theta)$ is the density corresponding to the initial observation X_1. Assuming that f is differentiable we have

$$\frac{d \log L_n}{d\theta} \simeq \sum_{k=1}^{n-1} \frac{d \log f(x_k, x_{k+1}; \theta)}{d\theta} = \sum_{i=1}^{n-1} g(x_i, x_{i+1}; \theta), \quad \text{say},$$

since the first term can be neglected under the regularity conditions. We may now impose the following regularity conditions:

(RC.1′), (RC.2′), (RC.3′)

same as (RC.1), (RC.2) and (RC.3) respectively with $f(x; \theta)$ replaced by $f(x, y; \theta)$ and E_θ being taken as the expectation corresponding the stationary distribution. In addition to these assumptions we need (RC.4). For each $x \in \mathcal{X}$ and $A \in \mathcal{F}$,

$$\pi(A; \theta) = 0 \Rightarrow p(x, A; \theta) = 0.$$

The last condition precludes transient states and is needed for the validity of the law of large numbers.

Using the above conditions Billingsley (1961) has shown that the limit properties (i) to (iv) mentioned earlier for the classical model also hold for the Markov processes under consideration. The proofs of Billingsley (1961) also hold for the classical model by taking $f(x, y; \theta) = f(x; \theta)$, the condition (RC.4) being satisfied trivially.

The basic technique for establishing the asymptotic properties of the ML estimators for both the classical model and Markov processes involves an expansion of $d \log L_n/d\theta$ in θ by the mean value theorem and an application of the law of large numbers (LLN) and the central limit theorem (CLT) on the summands occurring in the expansion. Such a technique can of course be applied to any other dependent sequence of random variables provided an appropriate LLN and CLT can be proved for the process involved. Since under mild regularity conditions $d \log L_n/d\theta$ is in general a zero-mean Martingale and since the LLN and CLT are well known for Martingales (under appropriate regularity conditions) it is clear that the properties of the ML estimator continue to remain valid for more general processes (ergodic type, see Section 5). See, for instance, Basawa, Feigin, and Heyde (1976) for further details.

5. A general model for discrete-time processes

5.1. The model

Let $\{X_k, k \geq 1\}$ be a discrete-time process defined on a probability space $(\mathscr{X}, \mathscr{F}, P_\theta)$, $\theta \in \Theta$, an open subset of the k-dimensional Euclidean space. Suppose the vector $X(n) = (X_1, X_2, \ldots, X_n)$ takes values in the measure space $(\mathscr{X}^n, \mathscr{F}^n)$ and possesses a density $p(x(n); \theta)$ with respect to a σ-finite product measure μ^n. The density $p(x(n); \theta)$ viewed as a function of θ is a likelihood function as remarked earlier. For simplicity we retain the same notation $p(\cdot\,; \theta)$ to denote the likelihood function.

Denote the *log-likelihood ratio* by Λ_n, where

$$\Lambda_n(\theta_n, \theta) = \log\{p(X(n); \theta_n)/p(X(n); \theta)\} \tag{5.1}$$

with $\theta_n = \theta + I_n^{-1/2}(\theta)h$, h being a $(k \times 1)$ vector of finite real numbers $h_1, h_2, \ldots, h_k$, and $I_n(\theta)$, a diagonal, non-random matrix with diagonal elements $a_{ni}(\theta)$, $i = 1, 2, \ldots, k$, such that for each $i = 1, 2, \ldots, k$, $0 < a_{ni}(\theta) < \infty$, $a_{ni}(\theta) \uparrow \infty$ as $n \uparrow \infty$, and for all $\theta \in \Theta$. Thus, $\{\theta_n\}$ is a sequence of values such that $\theta_n \to \theta$ as $n \to \infty$. It is assumed that for all $n \geq 1$, $\theta_n \in \Theta$. The elements $a_{ni}(\theta)$ of $I_n(\theta)$ will be chosen such that Λ_n will have a certain asymptotic behavior. A large number of problems in the general area of asymptotic parametric inference can be reduced essentially to the problem of studying the limiting behavior of Λ_n in (5.1).

Let us suppose that $p(X(n); \theta)$ is differentiable in θ at least twice. Let $S_n(\theta)$ be the $(k \times 1)$ *score vector* with elements $\partial \log p/\partial\theta_i$, $i = 1, 2, \ldots, k$, and denote the matrix of second partial derivatives (with a minus sign) by $B_n(\theta)$. Thus

$$B_n(\theta) = -(\partial^2 \log p/\partial\theta_i\, \partial\theta_j), \quad i, j = 1, 2, \ldots, k.$$

We may then formalize our assumptions regarding the limiting behavior of Λ_n as follows:

For a suitable choice of $I_n(\theta)$, assume that (A)–(C) below hold:

(A) $\Lambda_n(\theta_n, \theta) = h^{\mathrm{T}}\Delta_n(\theta) - \frac{1}{2}h^{\mathrm{T}}G_n(\theta)h + \mathrm{o}_p(1)$, where $\Delta_n(\theta) = I_n^{-1/2}(\theta)S_n(\theta)$, $G_n(\theta) = I_n^{-1/2}(\theta)B_n(\theta)I_n^{-1/2}(\theta)$, $\mathrm{o}_p(1)$ denotes terms which converge to zero in probability as $n \to \infty$, under $P_{n,\theta}$ probability, and T the transpose.

(B) $\mathscr{L}(\Delta_n(\theta), G_n(\theta)) \to \mathscr{L}(\Delta(\theta), G(\theta))$ as $n \to \infty$, under $P_{n,\theta}$ probability, where $\mathscr{L}(X_n) \to \mathscr{L}(X)$ denotes convergence in distribution, $G(\theta)$ is a possibly random $(k \times k)$ non-negative definite matrix and $\Delta(\theta)$ is a random vector.

(C) $\Delta(\theta)$ has the same distribution as $G^{1/2}(\theta)Z$, where Z is a $(k \times 1)$ vector of independent and identically distributed (i.i.d.) N(0, 1) variables, and Z is independent of $G(\theta)$.

Assumption (A) above states that in the Taylor expansion of $\log p(X^n, \theta_n)$ around θ the third and higher order terms may be ignored for large n. If differentiation under the integral sign of $\log p$ is permitted it is easily verified that $\{S_n(\theta), n = 1, 2, \ldots\}$ is a zero-mean Martingale. The Cramér–Wold device and Martingale limit theorems are often useful in verifying conditions (B) and (C). Various sufficient conditions which essentially ensure (A) to (C) are available in the literature. See for example, Wald (1948), Silvey (1961), Bhat (1974), Basawa et al. (1976) and Weiss (1971, 1973, 1975). If $\{X_j, j = 1, 2, \ldots\}$ are i.i.d. random variables, and $a_{ni}(\theta) = n$, $(i = 1, \ldots, k)$ are the diagonal elements of $I_n(\theta)$ the above assumptions specify the 'locally asymptotically normal' (L.A.N.) model of Le Cam (see Hajek (1971)); in this special case the matrix $G(\theta)$ will be non-random. If $\{X_j\}$ form a Markov process on a general state space $\mathscr{X}$ the assumptions in Billingsley (1961) would imply (A) to (C) with $a_{ni}(\theta) = n$ and $G(\theta)$ non-random. The differentiability condition on $\log p$ may be relaxed and replaced by a weaker condition of differentiability in quadratic mean; this, for instance, is done in Roussas (1972) for Markov processes. The works mentioned so far require $G(\theta)$ to be non-random. Feigin (1975), Heyde (1978), Basawa and Scott (1977) and Basawa and Koul (1979) discuss specifically the case where $G(\theta)$ is a non-degenerate random variable; Basawa and Koul (1979) study the vector parameter model as formulated above. The distinction between the two cases: viz.

(a) $G(\theta)$, non-random and

(b) $G(\theta)$ non-degenerate random is rather important and for future reference let us call them *ergodic* and *non-ergodic* families respectively.

It turns out that the standard Fisher–Cramér–Wald–Rao type results (valid for the i.i.d. model) can be extended to the ergodic general model. In particular, for the ergodic model, with $k = 1$, the maximum likelihood estimator (M.L.E.) is typically asymptotically normal and is efficient in the sense of having a smallest asymptotic variance in the class of consistent asymptotically (uniformly) normal estimators. For the non-ergodic model, however the M.L.E. *does not* have a limiting normal distribution; instead its limiting distribution is weighted normal.

5.2. Some examples

Example 1 (A simple branching process). Let $\{Y_0 = 1, Y_1, Y_2, \ldots, Y_n\}$ denote successive generation sizes in a Galton–Watson branching process. Suppose Y_1 has the density

$$p_Y(y; \theta) = \theta^{-1}(1-\theta^{-1})^{y-1}, \quad y = 1, 2, \ldots, \quad (1 < \theta < \infty).$$

Thus, the mean and the variance of the 'offspring' distribution are

$$\mathrm{E}(Y_1) = \theta \quad \text{and} \quad \sigma^2(\theta) = \mathrm{Var}(Y_1) = \theta(\theta-1).$$

Here θ is an unknown parameter. Since $1 < \theta < \infty$ the process is super-critical, and also with probability 1, $Y_n \to \infty$ as $n \to \infty$. Using the notation of Section 5.1 we readily find that

$$S_n(\theta) = \sigma^{-2}(\theta) \sum_{k=1}^{n} (Y_k - \theta Y_{k-1}),$$

and

$$B_n(\theta) = \sigma^{-2}(\theta) \sum_1^n Y_{k-1} - \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\sigma^{-2}(\theta)\right) \sum_1^n (Y_k - \theta Y_{k-1}).$$

Suppose we choose

$$I_n(\theta)=\sigma^{-2}(\theta)(\theta^n-1)/(\theta-1).$$

It is not difficult to verify, via the Toeplitz lemma, and using the fact that $\sum_1^n(Y_k-\theta Y_{k-1})$ is a zero-mean Martingale that

$$G_n(\theta)=B_n(\theta)/I_n(\theta)\xrightarrow{p}G,$$

where G is the limit in probability of (Y_n/θ^n). It is well known for the present case that G is a positive non-degenerate random variable and has the exponential distribution with unit mean. Conditions (A) to (C) can now be easily verified (see e.g. Basawa and Scott (1976)) with $\Delta(\theta)$ having the distribution of $G^{1/2}Z$, i.e. double exponential distribution with density $e^{-\sqrt{2}|u|}/\sqrt{2}$, $-\infty<u<\infty$. The M.L.E. of θ is seen to be

$$\hat{\theta}_n=\sum_1^n Y_k\Big/\sum_1^n Y_{k-1}.$$

One can also verify that $I_n^{1/2}(\theta)(\hat{\theta}_n-\theta)$ is asymptotically distributed as the Student's t distribution with 2 degrees of freedom.

Example 2 (A regression model with autocorrelated errors). Let $\{Y_j, 1\leq j\leq n\}$ satisfy the regression equation

$$Y_j=\alpha x_j+Z_j,$$

where the errors Z_j in turn satisfy the autoregressive equation

$$Z_j=\beta Z_{j-1}+\varepsilon_j,\quad |\beta|>1$$

with $\{\varepsilon_j, 1\leq j\leq n\}$ i.i.d. N(0, 1) variables, and $\varepsilon_j=x_j=0$ for $j\leq 0$. The 'design' variables x_j are assumed fixed and satisfy the regularity conditions

(i) $\{(\max_{1\leq j\leq n} x_j^2)/(\sum_1^n x_i^2)\}\to 0$ and

(ii) $\{\sum_1^n x_ix_{i-1}/\sum_1^n x_i^2\}\to b$, $|b|<\infty$.

Here $\theta^{\mathrm{T}}=(\alpha,\beta)$ is an unknown parameter.

A generalization of this model is discussed by Basawa and Koul (1979) who show that the model belongs to the non-ergodic family and satisfies conditions (A) to (C). The matrix $G(\theta)$ is seen to be

$$G=\begin{pmatrix}1 & 0\\ 0 & W\end{pmatrix},$$

where W is a chi-square random variable with 1 degree of freedom. The normalizing matrix $I_n(\theta)$ used is

$$I_n(\theta)=\begin{pmatrix}a_{n1}(\theta) & 0\\ 0 & a_{n2}(\theta)\end{pmatrix},$$

where

$$a_{n1}(\theta)=\sum_1^n (x_i-\beta x_{i-1})^2 \quad\text{and}\quad a_{n2}(\theta)=(\beta^{2n}-1)/(\beta^2-1).$$

The limit distributions of the likelihood ratio statistic and the score-statistic for testing the composite hypothesis $H:\beta=\beta_0$ with α as a nuisance parameter are derived in Basawa and Koul (1979).

Note that if the autoregressive parameter β is such that $|\beta|<1$ (stationary case), the above choice of $a_{n1}(\theta)$ and $a_{n2}(\theta)=n$ gives

$$G=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

and then the model belongs to the ergodic family.

Example 3 (A moving average process). Let $\{Y_j, 1\leq j\leq n\}$ satisfy the moving average equation

$$Y_j=\varepsilon_j-\alpha\varepsilon_{j-1},$$

where $\{\varepsilon_i, 0\leq i\leq n\}$ are i.i.d. N$(\theta, 1)$ variables.

Assume that $|\alpha|<1$ and α known. θ is an unknown parameter. It can be verified (cf. Prasad (1973)) that this example belongs to the ergodic family with the choice $I_n(\theta)=2n$ and $G(\theta)\equiv 1$. The usual properties of the M.L.E. of θ and of the likelihood-ratio statistic follow readily.

Example 4 (Markov processes on a general state space). Let $\{Y_k, k=1,2,\ldots\}$ be a Markov process on a state space $\mathscr{X}$ with time-homogeneous transition measures

$$p(A\,|\,x;\theta)=P_\theta(Y_{k+1}\in A\,|\,Y_k=x),\quad x\in\mathscr{X},\ A\in\mathscr{F},$$

where $\mathscr{F}$ is the σ-field of subsets of $\mathscr{X}$.

$p(A\,|\,x;\theta)$ for each fixed $x\in\mathscr{X}$ and for each $\theta\in\Theta$ is a probability measure on $\mathscr{F}$. Θ is an open subset of the k-dimensional Euclidean space. Assume that there is a unique stationary probability measure $p_0(A;\theta)$ on $\mathscr{F}$ satisfying

$$p_0(A;\theta)=\int_{\mathscr{X}} p(A\,|\,x;\theta)p_0(\mathrm{d}x;\theta),\quad A\in\mathscr{F}.$$

Let μ be a σ-finite measure on $\mathscr{F}$ and assume that the transition densities $f(x,y;\theta)$ corresponding to the transition measures $p(A\,|\,x;\theta)$ exists w.r.t. μ satisfying

$$p(A\,|\,x;\theta)=\int_A f(x,y;\theta)\mu(\mathrm{d}y),\quad A\in\mathscr{F}.$$

The joint density of $Y^n=(Y_1,\ldots,Y_n)$ with respect μ^n assuming that the initial state $Y_0=y_0$ is fixed is then given by

$$P(y^n;\theta)=\prod_{k=1}^n f(y_{k-1},y_k;\theta).$$

Under standard assumptions (see Section 4.2) it is seen that this model belongs to the ergodic family with

$$I_n(\theta)=\begin{pmatrix} n & \cdots & 0 \\ \vdots & n & \vdots \\ 0 & \cdots & n \end{pmatrix} \quad \text{and} \quad G(\theta)=((\sigma_{ij}(\theta))),$$

where

$$\sigma_{ij}(\theta)=E_{p_0}\left\{\frac{\partial \log f(Y_{k-1}, Y_k;\theta)}{\partial\theta_i}\frac{\partial \log f(Y_{k-1}, Y_k;\theta)}{\partial\theta_j}\right\} \quad (i,j=1,\ldots k),$$

E_{p_0} denoting the expectation computed under the stationary distribution p_0. Note that $G(\theta)$ is the usual Fisher information matrix.

Properties of the M.L.E. and of the likelihood ratio tests for θ are discussed by Billingsley (1961).

Example 5 (A stationary ϕ-mixing process). Suppose $\{Y_k, k=1,2,\ldots\}$ is a ϕ-mixing process, i.e it satisfies

$$|\mathbf{P}(B\,|\,\mathcal{B}_{1,t})-\mathbf{P}(B)|\leq\phi_n$$

with probability 1, for all events $B\in\mathcal{B}_{(t+n,\infty)}$, for each $t\in(1,2,\ldots)$ and $n\in(1,2,\ldots)$ where $\mathcal{B}_{a,b}$ denotes the σ-field generated by $(Y_{i_1},\ldots,Y_{i_k})$, $a-1<i_1<i_2<\cdots<i_k<b+1$. $\{\phi_n\}$ is a sequence of positive numbers converging to zero as $n\to\infty$.

Prasad (1973) discusses the following example: Let $\boldsymbol{Y}^n$ be a $N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$ vector where

$$\boldsymbol{\mu}^{\mathrm{T}}=(\theta,\theta,\ldots,\theta) \quad \text{and} \quad \boldsymbol{\Sigma}=((\sigma_{ij})) \quad (i,j=1,\ldots,n),$$

where $\sigma_{ij}=\rho^{|i-j|}$, with $|\rho|<1$, ρ known. It can be verified that the process $\{Y_1, Y_2,\ldots\}$ is stationary ϕ-mixing with $\phi_n=\rho^n$. In this example θ is the only unknown parameter. It is seen that the example belongs to the ergodic family with $I_n(\theta)=n$ and $G(\theta)=(1-\rho)/(1+\rho)$.

Example 6 (A classical mixture experiment). Conditionally on $V=\nu$, suppose $Y_1, Y_2,\ldots,Y_n$ are i.i.d. $\mathrm{N}(\theta,\nu)$ variables. Unconditionally $(Y_1, Y_2,\ldots)$ form an exchangeable process and Y's are no longer independent. Let us assume, for instance, that V^{-1} is an exponential random variable with mean unity. It can be verified that the unconditional density of $\boldsymbol{Y}^n=(Y_1,\ldots,Y_n)$ belongs to the non-ergodic family with the choice $I_n(\theta)=n$ and $G(\theta)=V^{-1}$. It is interesting to note that the limit distributions of the M.L.E. of θ and of the likelihood-ratio statistic for this example are formally the same as those for Example 1.

6. Optimality results for ergodic models

6.1. Efficient estimation

For simplicity of presentation we assume θ to be one-dimensional. Assume the conditions (A) to (C) of Section 5 are satisfied with

$$B_n(\theta) = -\mathrm{d}^2 \log p/\mathrm{d}\theta^2 \quad \text{and} \quad I_n(\theta) = \mathrm{E}\{B_n(\theta)\}.$$

It is then seen that $G_n(\theta) \to 1$ by the ergodic theorem. Thus, in this section we can take the limiting $G(\theta) \equiv 1$. Our ergodic model is then characterized by the conditions:

(A′) $\Lambda_n(\theta_n, \theta) = h\Delta_n(\theta) - \frac{1}{2}h^2 + o_p(1)$, where h is a real number (non-zero), and

(B′) $\Delta_n(\theta) \Rightarrow \mathrm{N}(0, 1)$.

Let T_n^0 be a consistent estimator satisfying

$$\Delta_n(\theta) - I_n^{1/2}(\theta)(T_n^0 - \theta) \to 0, \quad \text{in probability,} \tag{6.1}$$

when θ is the true parameter. (6.1) implies that

$$I_n^{1/2}(\theta)(T_n^0 - \theta) \Rightarrow \mathrm{N}(0, 1). \tag{6.2}$$

Suppose T_n is any other consistent estimator such that

$$I_n^{1/2}(\theta)(T_n - \theta) \Rightarrow L, \tag{6.3}$$

where L has a continuous distribution F which is symmetric about zero (not necessarily normal). Following Weiss and Wolfowitz (1966) it is easily seen, under regularity conditions that any estimator T_n^0 satisfying (6.1) is efficient (with respect to the class of estimators satisfying (6.3) according to the Weiss–Wolfowitz criterion given in Section 3. In particular, if L is $\mathrm{N}(0, \sigma_T^2(\theta))$ the Weiss–Wolfowitz criterion reduces to the result

$$\sigma_T^2(\theta) \geq 1 \tag{6.4}$$

which establishes the classical asymptotic variance optimality for T_n^0 (see Section 3).

Proceeding as in Rao (1973) it is not difficult to verify that the ML estimator $\hat{\theta}_n$ satisfies (6.1) and hence one can conclude that $\hat{\theta}_n$ is optimal according to both the asymptotic variance and the more general Weiss–Wolfowitz criteria.

It would be of interest to obtain results regarding the rate of convergence of $I_n^{1/2}(\theta)(\hat{\theta}_n - \theta)$ to $\mathrm{N}(0, 1)$ where $\hat{\theta}_n$ is an ML estimator. For the special case when $\{X_k, k \geq 1\}$ is a stationary markov process and for the case when $\{X_k\}$ is a sequence of independent, not necessarily identically distributed random variables, Prakasa Rao (1973, 1975) has shown that for every compact set $K \subset \Theta$, there exists a constant C_K such that for all $n \geq 1$ and $t \in R$,

$$\sup_{\theta \in K} |P_{n,\theta}(I_n^{1/2}(\theta)(\hat{\theta}_n - \theta) < t) - \Phi(t)| \leq C_K n^{-1/2}, \tag{6.5}$$

where $I_n(\theta) = na(\theta)$, and $\Phi(t)$ is the distribution function of $\mathrm{N}(0, 1)$.

6.2. Efficient tests

Consider the model satisfying (A) to (C) of Section 5 with $I_n(\theta)$ having diagonal elements all equal to n and $G(\theta)$ a non-random matrix with rank $l \leq k$. It is assumed that θ is k-dimensional. Consider the problem of testing the hypothesis $H: \theta = \theta_0$. Let $\{\theta_n\}$ be a sequence of alternative values of θ, where $\theta_n = \theta_0 + I_n^{-1/2}\delta_n$, δ_n being a $(k \times 1)$ vector of real numbers such that $\delta_n \to \delta$ as $n \to \infty$. It follows from assumption (C) that

$$\mathscr{L}(\Delta_n(\theta_0)) \to \mathrm{N}_k(0, G(\theta_0)), \quad \text{under } P_{n,\theta_0} \text{ probability}, \tag{6.6}$$

where $\mathscr{L}(Y_n) \to \mathscr{L}(Y)$ denotes the distribution convergence of Y_n, and $\mathrm{N}_k(A, B)$ is the k-variate normal distribution with mean vector A and covariance matrix B. Under assumptions (A) to (C) it can be verified that the sequences of probability measures $\{P_{n,\theta_0}\}$ and $\{P_{n,\theta_n}\}$ are mutually contiguous (via condition (S_3) of Roussas (1972), p. 11). This fact enables us to compute the limiting distribution of $\Delta_n(\theta_0)$ under the alternative $\{P_{n,\theta_n}\}$ probability (see e.g. Roussas (1972), p. 54, Theorem 4.6 for the special case of Markov processes). Thus,

$$\mathscr{L}(\Delta_n(\theta_0)) \to \mathrm{N}_k(G(\theta_0)\delta, G(\theta_0)), \quad \text{under } P_{n,\theta_n} \text{ probability}. \tag{6.7}$$

One may consider the test statistic

$$Q_n = \Delta_n^{\mathrm{T}}(\theta_0) G^{-}(\theta_0) \Delta_n(\theta_0), \tag{6.8}$$

where $G^{-}(\theta_0)$ is any generalized inverse of the matrix $G(\theta_0)$. It follows from (6.6) and (6.7) that

$$\mathscr{L}(Q_n) \to \begin{cases} \text{chi-square } (l) \text{ under } P_{n,\theta_0} \text{ probability}, \\ \text{Non-central chi-square } (l, \lambda) \text{ under } P_{n,\theta_n} \text{ probability}, \end{cases} \tag{6.9}$$

where l = degrees of freedom = rank of $G(\theta_0)$ and λ = non-centrality parameter = $\delta^{\mathrm{T}} G(\theta_0) G^{-}(\theta_0) G(\theta_0)\delta = \delta^{\mathrm{T}} G(\theta_0)\delta$.

Eq. (6.9) gives us the necessary limit distributions to derive the size and the power of the test based on Q_n. Note that the statistic Q_n requires the knowledge of the limiting matrix $G(\theta_0)$ which in the present case turns out to be the Fisher information matrix per observation and is usually easy to compute.

Alternatively, one may consider the likelihood-ratio statistic, $-2 \log \lambda_n$, where

$$\lambda_n = p(\boldsymbol{Y}^n; \theta_0)/p(\boldsymbol{Y}^n; \hat{\theta}_n), \tag{6.11}$$

$\hat{\theta}_n$ being the M.L.E. of θ. It can be shown that

$$(-2 \log \lambda_n - Q_n) \xrightarrow{p} 0 \tag{6.12}$$

both under P_{n,θ_0} and P_{n,θ_n} probabilities. Thus, the likelihood ratio statistic $-2 \log \lambda_n$ is asymptotically equivalent to the score statistic Q_n. If $G(\theta_0)$ is not known or not easy to compute one may use $-2 \log \lambda_n$ rather than Q_n since the former statistic does not require the knowledge of $G(\theta_0)$ (but requires computation of $\hat{\theta}_n$).

It is not difficult to show, using standard techniques that both the statistics discussed above are optimal in the sense of maximizing Pitman power (i.e. the limit of the power function at θ_n), among all tests of the same size. (See Section 3 for definitions.) The limit distribution of $-2\log\lambda_n$ and Q_n for testing a composite hypothesis involving some nuisance parameters can easily be obtained. See Dzhaparidge (1977) for details.

Large-sample tests based on the log-likelihood were also studied by Prakasa Rao (1974) and optimal asymptotic tests in a certain sense were discussed by Bhat and Kulkarni (1972) for the ergodic type models.

7. Optimality results for non-ergodic models

7.1. Estimation

Consider the model of Section 5 (satisfying assumptions (A) to (C) therein) and for simplicity let θ be one-dimensional. Assume $G(\theta)$ to be a non-degenerate positive random variable. In particular, we have

$$\mathscr{L}(\Delta_n(\theta))\to\mathscr{L}(G^{1/2}(\theta)Z), \tag{7.1}$$

where Z is a N(0, 1) random variable independent of $G(\theta)$. The limiting distribution of $\Delta_n(\theta)$ is thus a weighted normal distribution and is *not* normal.

Let T_n^0 be any consistent estimator satisfying

$$I_n^{1/2}(\theta)(T_n^0-\theta)-G^{-1}(\theta)\Delta_n(\theta)\to 0, \quad \text{in probability} \tag{7.2}$$

when θ is the true parameter. It then follows that

$$\mathscr{L}(I_n^{1/2}(\theta)(T_n^0-\theta))\to\mathscr{L}(G^{-1/2}(\theta)Z), \tag{7.3}$$

when θ is the true parameter. The limiting distribution of T_n^0 is thus non-normal. Let T_n be any other consistent estimator such that

$$\mathscr{L}(I_n^{1/2}(\theta)(T_n-\theta)\to\mathscr{L}(T), \tag{7.4}$$

where T has some continuous distribution symmetric at the origin. Following the arguments of Weiss and Wolfowitz (1966) it is easily seen that T_n^0 is optimal according to the Weiss–Wolfowitz criterion in Section 3 with respect to the class of estimators satisfying (7.4). This fact was verified rigorously by Heyde (1978) for the ML estimator $\hat\theta_n$. Basawa and Scott (1977) have shown that the ML estimator $\hat\theta_n$ satisfies (7.2).

Since the limit distribution of the ML estimator $\hat\theta_n$ is not normal the asymptotic variance criterion of Section 3 is not relevant here. In fact, for the branching process example discussed in Example 1 it is known that $I_n^{1/2}(\theta)(\hat\theta_n-\theta)$ converges to a random variable whose variance is infinite.

See Heyde (1975), and Basawa and Scott (1979) for a detailed discussion of a criterion related to (7.2).

7.2. Asymptotic tests

Consider the problem of testing $H: \theta=\theta_0$ against the one sided alternative $K: \theta>\theta$, where θ is one-dimensional. Basawa and Scott (1977) show that the statistic $\Delta_n(\theta_0)$ is optimal according to the local power criterion (see Section 3). However, Sweeting (1978) and Feigin (1978) showed, in particular non-ergodic models (branching process, and conditional exponential family respectively), that $\Delta_n(\theta_0)$ is *not* optimal according to the Pitman criterion (see Section 3). In fact, it is possible to show that a modified version of the score-statistic $\Delta_n^*(\theta_0)$ is optimal according to the Pitman criterion, where

$$\Delta_n^*(\theta_0)=\Delta_n(\theta_0)-\tfrac{1}{2}hG_n(\theta_0) \tag{7.5}$$

with

$$G_n(\theta_0)=I_n^{-1}(\theta_0)B_n(\theta_0) \quad \text{and} \quad B_n(\theta_0)=(-\mathrm{d}^2\log p/\mathrm{d}\theta^2)_{\theta_0}.$$

The statistic $\Delta_n^*(\theta_0)$ defined in (7.5) is seen to satisfy the Pitman criterion at h corresponding to the sequence of alternatives $\theta_n=\theta_0+I_n^{-1/2}(\theta_0)h$, $h>0$.

Basawa and Koul (1980) consider the vector parameter case and the problem of testing a composite hypothesis involving nuisance parameters and show that a modified score-statistic analogous to the one defined in (7.5) is asymptotically minimax in the sense of Weiss and Wolfowitz (1969).

Basawa and Koul (1979) study the asymptotic properties of the likelihood-ratio test for testing a composite hypothesis involving several parameters and obtain non-standard results. For the special case of a single parameter and testing a simple hypothesis it turns out that the likelihood-ration statistic is asymptotically equivalent to

$$T_n=\{B_n(\hat\theta_n)(\hat\theta_n-\theta_0)\}^2,$$

where $\hat\theta_n$ is the ML estimator. Under $H: \theta=\theta_0$, the limit distribution T_n is chi-square with one degree of freedom. The statistic T_n is therefore simpler to use. However, T_n is not optimal according to either the local or the Pitman criterion. This fact is to be contrasted with the corresponding result for the ergodic model of Section 6 for which T_n and $\Delta_n^2(\theta_0)$ are asymptotically equivalent and both statistics are optimal according to both the local and the Pitman criterion. Since in the ergodic model $G_n(\theta)$ converges to a constant, the second term in $\Delta_n^*(\theta_0)$ defined by (7.5) can be ignored.

8. Continuous time processes

A basic and fundamental work in the area of inference for continuous time processes is due to Grenander (1950). Billingsley (1961) and Ranneby (1975) considered continuous-time Markov chains. Akritas (1978) studied estimation problems for Markov processes, semi-Markov processes and related jump-type processes in continuous time. Feigin (1976) surveyed estimation problems for non-ergodic type continuous-time processes including diffusion processes. Athreya and Keiding (1977) studied estimation in continuous-time branching processes. Taraskin (1974), Brown and Hewitt (1975) Le Breton (1975) and Kutoyans (1977) among others investigated various inference problems for diffusion processes. The monograph by Cox and Lewis (1966) gives a survey of early work on point processes. See Basawa and Prakasa Rao (1980) for further references.

8.1. Jump-type processes

8.1.1. The likelihood function

Let $\{Y_u, u \geq 0\}$ be a continuous-time process with the state space $\mathscr{X} \subset \mathbf{R}$. Denote by $\mathscr{F}_t = \sigma\{Y_u, 0 \leq u \leq t\}$ the σ-field generated by the family of random variables $\{Y_u, 0 \leq u \leq t\}$. P_θ^t denotes the probability measure defined on $\mathscr{F}_t$, where $\theta \in \Theta \subset \mathbf{R}^k$ (k-dimensional Euclidean space). It is assumed that for every pair θ_1, $\theta_2 \in \Theta$ the probability measures $P_{\theta_1}^\infty$ and $P_{\theta_2}^\infty$ are mutually absolutely continuous. Define the density (or the likelihood function) corresponding to the family of random variables $\{Y_u, 0 \leq u \leq t\}$ as a Radon–Nikodym derivative of P_θ^t with respect to $P_{\theta_0}^t$, where θ_0 is some fixed value of θ. Thus,

$$p(y_u, 0 \leq u \leq t; \theta) = \frac{\mathrm{d}P_\theta^t}{\mathrm{d}P_{\theta_0}^t}(y_u, 0 \leq u \leq t).$$

The calculation of a Radon–Nikodym derivative for continuous-time processes is not always easy. If the process is of a purely jump-type, however in a majority of cases, the density for the continuous observation over the interval $(0, t)$ can be deduced (at least heuristically) simply by concentrating on the jump-size distribution and the distribution for between-jump intervals, provided that only a finite number of jumps occur with probability one in any finite interval. In this section we shall adopt this heuristic derivation of the density and confine ourselves, from now on, to the purely discontinuous process. Diffusion processes will be discussed later on.

At this stage one may formulate the ergodic and non-ergodic models for $p(y_u, 0 \leq u \leq t; \theta)$ on the same lines as for the discrete-time case ((A) to (C) in Section 5.1) the limits being for $t \to \infty$ rather than as $n \to \infty$. Similar discussion as in the previous sections, will apply for the continuous-time models. We now turn to examples.

8.1.2. Some Examples

Example 7 (Stable process). Let $\{Y_u, u \geq 0\}$ be an additive process with non-decreasing sample paths. In particular, suppose the increments in any disjoint time-intervals are independent and have a positive stable distribution. Let Y_t, for any $t > 0$, have a distribution with Laplace transform

$$\mathrm{E}[\exp\{-\lambda Y_t\}] = \exp\{-t\alpha\lambda^\beta / \Gamma(1-\beta)\} \quad (\alpha > 0, 0 < \beta < 1).$$

Such a process was used by Brockwell and Chung (1975) as an input to a dam model. $\theta^{\mathrm{T}} = (\alpha, \beta)$ is the parameter of interest. It is well known, that, for the stable process under consideration, an infinite number of jumps occur in any finite interval. We, therefore, consider instead a related process where only jumps of size $\geq \varepsilon$ are recorded, where $\varepsilon > 0$ is a predetermined small number. Let $N(t, \varepsilon)$ denote the number of jumps in $(0, t)$ whose size $\geq \varepsilon$, $\{U_k(\varepsilon), k = 1, 2, \ldots\}$ the sequence of jumps of size $\geq \varepsilon$, and $\{Z_k; k = 1, 2, \ldots\}$ the sequence of time points at which such jumps occur. $\{(Z_k, U_k(\varepsilon)), k = 1, 2, \ldots\}$ will then determine a compound Poisson process. We can now construct the likelihood function based on the realization

$$\{(Z_k, U_k(\varepsilon)), k = 1, 2, \ldots, N(t, \varepsilon)\},$$

and this is given by

$$L(\alpha, \beta) = e^{-t\psi} \prod_{i=1}^{N} \lambda(U_i),$$

where $\psi = \int_\varepsilon^\infty \alpha\beta u^{-1-\beta}\, du$, $\lambda(u) = \alpha\beta u^{-1-\beta}$, U_i and N are abbreviations for $U_i(\varepsilon)$ and $N(t, \varepsilon)$ respectively. It is clear from the work in Basawa and Brockwell (1978) that this model belongs to the ergodic family with the diagonal elements of $I_t(\theta)$ being equal to t and $G(\theta)$ being the Fisher information matrix with elements

$$G_{11}(\theta) = [\beta^{-2} + \{\log \varepsilon + \Gamma'(1-\beta)/\Gamma(1-\beta)\}^2]\psi, \qquad G_{22}(\theta) = \psi/\alpha^2,$$

and

$$G_{12}(\theta) = G_{21}(\theta) = \{\log \varepsilon + \Gamma'(1-\beta)/\Gamma(1-\beta)\}^2\psi.$$

Standard asymptotic results (cf. Section 6.1 and 6.2) regarding the M.L.E. and likelihood ratio tests apply as $t \to \infty$, for any fixed ε. Basawa and Brockwell (1978, 1980) discuss the limiting behavior of the M.L.E.'s of α and β for fixed t and as $\varepsilon \to 0$, which may be of independent interest. See Basawa (1980) for the properties of a conditional test of β.

Example 8 (Markov processes on continuous time). Let $\{Y_u, u \geq 0\}$ be a Markov process on the state space $\mathscr{X}$, a Borel subset of a Euclidean space, and transition measures

$$p(t, y, A; \theta) = \mathbf{P}\{Y_{u+t} \in A \mid Y_u = y; \theta\}, \quad A \in \mathscr{F},$$

where $\theta \in \Theta \subset E^k$. It is assumed that the sample functions of the process are right continuous step-functions. Furthermore, suppose

$$\lim_{t \to 0} p(t, y, \{y\}; \theta) = 1 \quad \text{for all } y \text{ and } \theta.$$

Consider the discrete-time imbedded process $\{(X_k, Z_k), k = 0, 1, 2, \ldots\}$ where X_k are the successive distinct states of the process and Z_k are the times for which the process stays in the states X_k. The imbedded process is a Markov process (on discrete time) on the state space $\mathscr{X} \times R$, $R = (0, \infty)$.

Suitable assumptions on the transition densities of the discrete-time Markov process $\{(X_k, Z_k), k = 0, 1, 2, \ldots\}$ (see Billingsley (1961)) ensure that results analogous to Example 4 apply. See Billingsley (1961) for details, and for specific applications. (Also, see Example 9 below.)

Example 9 (A queueing process). This is an application of Example 8. Let λ_j and μ_j denote the arrival rate and the service rate given that the queue-size is j. Assume that the inter-arrival times and the service times are exponential random variables. Set $\mu_0 = 0$, suppose λ_j for $j \geq 1$ are known functions of an unknown parameter $\theta \in \Theta \subset E^k$. Also, μ_j for $j \geq 0$ are assumed to be known functions of θ. The evolution of this queueing process follows that of a birth and death process which : a continuous time Markov process on the state space $\mathscr{X} = \{0, 1, 2, \ldots\}$. The likelihood function corresponding to the realization of the process over $(0, t)$ can easily be deduced using the imbedded process, and is given by (assuming a fixed initial state)

$$\left(\prod_{j=0}^{\infty} \lambda_j^{u_j}(\theta)\right)\left(\prod_{j=1}^{\infty} \mu_j^{d_j}(\theta)\right)\left(\prod_{j=0}^{\infty} \exp(-(\lambda_j(\theta) + \mu_j(\theta)))\nu_j\right),$$

where

$$u_j = \text{frequency of transitions } j \to j+1 \text{ in } (0, t),$$

$$d_j = \text{frequency of transitions } j \to j-1 \text{ in } (0, t),$$

$$\nu_j = \text{total time spent, during } (0, t), \text{ in the state } j.$$

Assuming that the stationary distribution for the queue-size process $\{Y_u, u \geq 0\}$ exists it can be verified that this model belongs to the ergodic family. See Wolff (1965) for the properties of the M.L.E. of θ and of likelihood ratio tests.

Example 10 (A semi-Markov process). Let $\{Y_u, u \geq 0\}$ be a jump-type process having an imbedded process $\{(X_k, Z_k), k = 0, 1, 2, \ldots\}$, where, as in Example 8, $\{X_k\}$ are the successive distinct states the process passes through, and Z_k are the times spent in the states X_k. Assume that $\{(X_k, Z_k), k = 0, 1, 2, \ldots\}$ is a Markov process on $\mathscr{X} \times R$, $R = (0, \infty)$. In particular, suppose $\{X_k\}$ is a Markov chain on $\mathscr{X} = \{1, 2, \ldots, m\}$, and conditional on $\{X_k\}$, Z_k are independent gamma variates with index x_k and mean $x_k\theta$. The process $\{Y_u, u \geq 0\}$ is then not Markovian, but it is a semi-Markov process. (For $\{Y_u, u \geq 0\}$ to be Markovian one must assume that Z_k are independent exponential (given X_k) random variables.) Suppose, for simplicity, that the transition probabilities of the Markov chain $\{X_k\}$ do not depend on θ. The problem is that of drawing inferences about θ from a realization $\{Y_u, 0 \leq u \leq t\}$. Although $\{Y_u, u \geq 0\}$ is not Markovian, we can easily construct the likelihood function using the imbedded process $\{(X_k, Z_k), k = 1, 2, \ldots\}$ which is Markovian on discrete time. The problem then reduces to that of Example 4. See Basawa (1974), and Feigin (1976) for details on the present example. This example is seen to belong to the ergodic family with the choice $I_t(\theta) = t$.

Example 11 (Pure birth process). Let $\{Y_u, u \geq 0\}$ be a linear birth process with birth rate θ. This is a Markov process on $\mathscr{X} = \{1, 2, \ldots\}$, with the imbedded process (X_k, Z_k), $k = 0, 1, 2, \ldots$, where $X_k = k$ (assuming $Y_0 = 1$) with probability 1, and Z_k are independent exponential random variables with rates $\theta(k+1)$. The likelihood function of $\{Y_u, 0 \leq u \leq t\}$ using the imbedded process, is easily constructed and is given by (ignoring terms free from θ)

$$e^{-\theta S_t} \theta^{B_t},$$

where $S_t = \int_0^t Y_u \, du$, and B_t is the number of births in $(0, t)$. The limit distribution of the M.L.E. of θ was derived by Keiding (1974). It is apparent from Keiding's work that this model belongs to the non-ergodic family with $I_t(\theta) = (e^{\theta t} - 1)/\theta^2$, and $G(\theta) \stackrel{d}{=} V$, where V is an exponential random variable with mean unity. If we observe $\{Y_u\}$ at discrete equidistant points, say $k = 1, 2, \ldots$, the resulting discrete-time process will be a super-critical branching process with a geometric offspring distribution discussed in Example 1. Thus, Example 11 is a continuous time analogue of Example 1.

8.2. Diffusion processes

8.2.1. The likelihood function

A diffusion process $\{Y_u, u \geq 0\}$ is a process on continuous time with continuous sample paths, and satisfying certain further conditions. We may view the process $\{Y_u\}$ as a solution of the stochastic differential equation

$$dY_u = \mu(Y_u) \, du + \sigma(Y_u) \, dW_u, \quad u \geq 0,$$

where $\mu(\cdot)$ and $\sigma(\cdot) \geq 0$ are continuous functions and $\{W_u, u \geq 0\}$ is a standard Wiener process. (See Gikhman and Skorokhod (1972).) Suppose that the function $\mu(Y_u)$ is a known function of an unknown parameter θ, and wirte $\mu(Y_u; \theta)$. From now on we consider the equation

$$dY_u = \mu(Y_u; \theta) \, du + dW_u, \quad u \geq 0,$$

where we have assumed, for simplicity, that $\sigma(\cdot) = 1$.

The problem is that of drawing inferences about θ from a continuous observation $\{Y_u; 0 \leq u \leq t\}$. Let $P_Y^t(\theta)$ and P_W^t denote the measures induced by $\{Y_u, 0 \leq u \leq t\}$ and $\{W_u, 0 \leq u \leq t\}$ respectively. Assume that

$$P_\theta\left\{\int_0^t \mu^2(Y_u; \theta) \, du < \infty\right\} = 1, \quad \forall \theta \in \Theta.$$

Then, $P_Y^t(\theta)$ is absolutely continuous with respect to P_W^t and it can be shown (see Kailath and Zakai (1971), or Lipcer and Sirjaev (1972)) that the Radon–Nikodym derivative of $P_Y^t(\theta)$ with respect to P_W^t is given by

$$\frac{dP_y^t(\theta)}{dP_w^t}(Y_u, 0 \leq u \leq t) = \exp\left\{\int_0^t \mu(y_u; \theta) \, dY_u - \tfrac{1}{2} \int_0^t \mu^2(Y_u; \theta) \, du\right\}.$$

Since

$$\mathrm{d}P_Y^t(\theta)/\mathrm{d}P_Y^t(\theta_0) = \left\{\frac{\mathrm{d}P_Y^t(\theta)}{\mathrm{d}P_W^t} \Big/ \frac{\mathrm{d}P_Y^t(\theta_0)}{\mathrm{d}P_W^t}\right\}, \quad \theta_0 \text{ fixed},$$

it follows that the likelihood function ignoring terms free from θ is given by

$$\exp\left\{\int_0^t \mu(Y_u;\theta)\,\mathrm{d}Y_u - \tfrac{1}{2}\int_0^t \mu^2(Y_u;\theta)\,\mathrm{d}u\right\},$$

where the first integral in the exponent is a stochastic integral. We now consider below a special case.

Example 12 (Ornstein–Uhlenbeck process). Let $\{Y_u, u \geqslant 0\}$ be a solution of

$$\mathrm{d}Y_u = \theta Y_u\,\mathrm{d}u + \mathrm{d}W_u, \quad u \geqslant 0, \quad Y_0 = 0.$$

Here $\mu(Y;\theta) = \theta Y_u$, and $\sigma(\cdot) = 1$.

The likelihood function is seen to be

$$\exp\left\{\theta\int_0^t Y_u\,\mathrm{d}Y_u - \tfrac{1}{2}\theta^2\int_0^t Y_u^2\,\mathrm{d}u\right\}.$$

For the explosive case, $\theta > 0$, it can be shown that this model belongs to the non-ergodic family with $I_t(\theta) = \mathrm{e}^{2\theta t}/2\theta$ and $G(\theta) \stackrel{d}{=} V$ where V is a χ_1^2 random variable. For the stationary case, $\theta < 0$, the example belongs to the ergodic family with the choice $I_t(\theta) = t$ and $G(\theta) = -1/2\theta$.

See Feigin (1976) and Brown and Hewitt (1975) for details on the limit distribution of the M.L.E. of θ for the cases $\theta > 0$ and < 0 respectively. See Dorogovcev (1974), Prakasa Rao (1979), and Prakasa Rao and Rubin (1979) for work on least-square estimation for diffusion processes.

9. Bayesian methods

9.1. *The Bernstein–Von Mises theorem*

A key result in the asymptotic theory of Bayesian inference is the so-called Bernstein–Von Mises theorem which states that the posterior distribution approaches the normal distribution in the mean as the sample size increases.

Suppose that $\{Y_n, n \geqslant 1\}$ is an arbitrary discrete parameter stochastic process defined on a probability space $(\Omega, \mathscr{F}, P_\theta)$, $\theta \in \Theta$ an open interval of R. Let $\theta_0 \in \Theta$ be the true parameter. Let Λ be a prior measure on $(\Theta, \mathscr{B})$, $\mathscr{B}$ the σ-algebra of Borel subsets of Θ. Suppose Λ has density $\lambda(\cdot)$ with respect to Lebesque measure and further suppose that $\lambda(\cdot)$ is continuous and positive in an open neighborhood of θ_0. Let $p_n(\theta \mid y(n)) = p_n(\theta \mid y_1, \ldots, y_n)$ be the posterior density of θ given the observations $y_1, \ldots, y_n$ corresponding to the prior probability density $\lambda(\cdot)$. ($p_n(\theta \mid y(n))$ is a conditional density of θ given $y(n)$.)

If Y_n is a discrete time stationary Markov process, it was shown that posterior density converges to normal in the mean by Borwanker, Kallianpur and Prakasa Rao (1971) using the methods of Bickel and Yahav (1969). Prasad (1973) (cf. Prasad and Prakasa Rao (1976) and Moore (1977)) proves this result for some special classes of processes. Let

$$p_n^*(t \mid y_1, \ldots, y_n) = n^{-1/2} p_n(\theta \mid y_1, \ldots, y_n),$$

where $t = n^{1/2}(\theta - \hat{\theta}_n)$, $\hat{\theta}_n$ being the ML estimator. Under some regularity conditions (cf. Basawa and Prakasa Rao (1980), Chapter 10, or Prakasa Rao (1974)), it can be shown that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \left| p_n^*(t|y(n)) - \left(\frac{i_0}{2\pi}\right)^{1/2} e^{-i_0 t^2/2} \right| dt = 0 \quad \text{almost surely,}$$

where i_0 is the limiting Fisher information. This result is known as the Bernstein–Von Mises theorem. As an application of the Bernstein–Von Mises theorem, one can obtain asymptotic properties of Bayes estimators T_n for a suitable class of loss functions (see Section 9.2) and show that the maximum likelihood estimators and Bayes estimators are asymptotically equivalent for smooth priors (cf. Prakasa Rao (1974)) extending a similar result of Borwanker, Kallianpur and Prakasa Rao (1971) in the Markov case.

9.2. Bayes estimation

As an application of the Bernstein–Von Mises theorem, one can obtain asymptotic properties of Bayes estimators. We define a Bayes estimator $T_n = T_n(y(n)) = T_n(y_1, \ldots, y_n)$ as an estimator which minimizes

$$B_n(\phi) = \int \tilde{W}(\theta, \phi) p_n(\theta|y(n))\, d\theta,$$

where $\tilde{W}(\theta, \phi)$ is a loss function on $\Theta \times \Theta$. Note that $B_n(\phi)$ is the posterior risk under the loss function $\tilde{W}(\cdot, \cdot)$ and prior density $\lambda(\cdot)$. Suppose such an estimator exists and is measurable. Further suppose that the loss function is of the type

$$\tilde{W}(\theta, \phi) = W(|\theta - \phi|),$$

where

$$W(x_1) \geq W(x_2) \quad \text{if } x_1 \geq x_2 \geq 0.$$

Let $\hat{\theta}_n$ be a MLE. It can be shown that

$$n^{1/2}(T_n - \hat{\theta}_n) \to 0 \quad \text{a.s.} \quad \text{as } n \to \infty,$$

under some further regularity conditions, (Prakasa Rao (1974)) extending earlier results of Borwanker, Kallianpur, Prakas Rao (1971). In particular, it follows that

$$n^{1/2}(T_n-\theta)\Rightarrow N(0, 1/I(\theta)) \quad \text{if } n^{1/2}(\hat{\theta}_n-\theta)\Rightarrow N(0, 1/I(\theta)).$$

It can also be shown that the asymptotic Bayes risk corresponding to the loss function W is the same as that of the maximum likelihood estimator. Recently (cf. Prakasa Rao (1977)), we have been able to show that

$$|T_n-\hat{\theta}_n|\leq C_K n^{-1}$$

uniformly over compact sets K of the parameter space giving an improved bound for discrete time stationary Markov processes generalizing similar results of Strasser (1977) in the i.i.d. case.

Generalized Bayes estimates for discrete time Markov chains including the transient Markov chains have been studied in Levit (1974) by using methods of Ibragimov and Khasminskii (1972). The peculiarity of the problem here consists in the fact that weighted probability laws arise as limit distributions of the estimator $\hat{\theta}_n$.

We mention that Doob (1949) obtained a fundamental result regarding consistency of Bayes estimators. He showed that Bayes estimators corresponding to a prior Λ are consistent except possibly on a set of Λ-measure zero under reasonable conditions. Schwartz (1965) proved consistency of Bayes estimators in the i.i.d. case under conditions weaker than those given in Le Cam (1953) or Wald (1949) for consistency of Bayes estimators and MLE's respectively. These results can be extended to arbitrary discrete time stochastic processes. Foutz (1974) discussed degenerate convergence of posterior distributions in the independent not necessarily identically distributed case. Similar results were obtained by Yamada (1976).

9.3. Bayesian testing

Let Λ be a prior measure on $(\Theta, \mathscr{B})$ which assigns positive probability to every non-empty open subset of Θ. Let the decision space D consist of two actions d_0 and d_1 such that for every θ at least one of $L(\theta, d_0)$ and $L(\theta, d_1)$ equals zero and both loss functions are non-negative and measurable in θ. Let

$$L(\theta)=\max\{L(\theta, d_1), L(\theta, d_0)\}.$$

Suppose that

$$\int_\Theta L(\theta)\Lambda(\mathrm{d}\theta)<\infty$$

and define

$$\Lambda^*(B)=\int_B L(\theta)\Lambda(\mathrm{d}\theta), \quad B\in\mathscr{B}.$$

Let $H=\{\theta: L(\theta, d_1)=L(\theta)\}$ and $H'=\Theta-H$.

A Bayes procedure for testing H versus H' is

$$\delta(y(n)) = \begin{cases} d_0, & \text{if } \int_H p(y(n); \theta)\Lambda^*(\mathrm{d}\theta) \geqslant \int_{H'} p(y(n); \theta)\Lambda^*(\mathrm{d}\theta), \\ d_1, & \text{otherwise} \end{cases}$$

and the Bayes posterior risk is

$$L_n = \min\left\{\frac{\int_H p(y(n); \theta)\Lambda^*(\mathrm{d}\theta)}{\int_\Theta p(y(n); \theta)\Lambda(\mathrm{d}\theta)}, \frac{\int_{H'} p(y(n); \theta)\Lambda^*(\mathrm{d}\theta)}{\int_\Theta p(y(n); \theta)\Lambda(\mathrm{d}\theta)}\right\}.$$

It can be shown that the posterior risk L_n goes to zero exponentially under some regularity conditions.

10. Nonparametric inference

10.1. Nonparametric estimation

In the previous sections, we studied statistical inference for stochastic processes under the assumption that the measures generated by the processes are completely known except for certain parameters which have to be estimated from the observed sample paths of the process. As in the classical situation, this might not always be possible and we may be forced to use nonparametric techniques for inference.

Suppose $\{X_n, n \geqslant 1\}$ is a strictly stationary stochastic process with a continuous one-dimensional marginal probability density $f(y)$. If the process consists of i.i.d. observations, then it was shown by Rosenblatt (1956) and Bickel and Lehmann (1969) in a more general case that there does not exist a function $h_n(y: x_1, \ldots, x_n)$ measurable in $(x_1, \ldots, x_n)$ for each y such that

$$E_f h_n(y; X_1, \ldots, X_n) = f(y)$$

for all y and for all continuous f. In other words, one can not obtain unbiased estimators of density under reasonable conditions. Borwanker (1967) showed that this negative result still holds when the observations are from a stationary process. It is natural to consider whether one can obtain estimators which are asymptotically unbiased, for instance. There are several methods of density estimation in the i.i.d. case. Most of these methods give estimators which, under comparable conditions, have similar properties as far as asymptotic behaviour, such as rates of mean square error, are concerned. The main methods which have been used in the dependent case are the kernel method of density estimation, the method using orthogonal expansions and the method of delta sequences. The kernel method has been used by Rosenblatt (1970) and Roussas (1968, 1969a, b) to obtain estimators of one-dimensional marginal and transition densities and the method of delta sequences by Prakasa Rao (1979) to obtain estimators of marginal densities for discrete time stationary Markov Processes. Delecroix (1975) used the method of orthogonal expansions for studying estimators of densities when the process is stationary and ϕ-mixing.

Banon (1978) discussed a nonparametric estimator for the marginal density of continuous time Markov processes. This estimator is a sequential analogue of the fixed sample density estimator using the Kernel method. Furthermore, it is recursive in nature, asymptotically unbiased and meansquare consistent at each x. A similar method can be adopted for estimation of the derivative of a density $p(x)$, when it exists. Banon (1978) has used these results for estimation of the drift coefficient $m(x)$ in a stochastic differential equation

$$\mathrm{d}\varepsilon_t = m(\varepsilon_t)\,\mathrm{d}t + \sigma(\varepsilon_t)\,\mathrm{d}W(t), \quad \varepsilon(0)=\varepsilon_0, \quad t \geqslant 0$$

by using the fact that $m(x)$, $\sigma(x)$ and $p(x)$ are related by equation

$$m(x) = \tfrac{1}{2}\{\sigma^{2\prime}(x) + \sigma^2(x)p'(x)/p(x)\}$$

for all x for which $p(x) \neq 0$. Recently we have obtained estimators of $p(\cdot)$ by using the method of delta families in Prakasa Rao (1979).

Ibragimov and Khasminskii (1977) studied the problem of estimating the value $F(S)$ of a given functional based on the observation of the solution $\{X(t), 0 \leqslant t \leqslant T\}$ of the stochastic differential equation of the type

$$\mathrm{d}X(t) = S(t) + \varepsilon\ \mathrm{d}W(t), \quad 0 \leqslant t \leqslant T.$$

We have discussed above one aspect of nonparametric estimation viz. estimation of the density when the observations are dependent and form either a continuous or discrete time process. Another aspect of the problem is to study the robustness properties of nonparametric estimators of location and scale for dependent data. Gastwirth and Rubin (1975b) investigated the effect of serial dependence in the data on the efficiency of robust estimators. They showed that if the observations are from a stationary process satisfying certain mixing condition then linear combinations of order statistics and the Hodges–Lehmann estimators are asymptotically normal. Related work on these problems is in Gastworth and Rubin (1971, 1975a) and Gastwirth, Rubin and Wolff (1967). Koul (1977) proved the asymptotic normality of the regression estimators when the errors in the regression model are stationary and strong mixing. A Chernoff–Savage representation for a general class of rank order statistics for stationary ϕ-mixing processes has been studied by Sen and Ghosh (1973). Sen (1977) has also discussed almost sure representation of linear combinations of order statistics with smooth weight functions from stationary ϕ-mixing processes. Portnoy (1977) obtained estimators of location parameters which are approximately asymptotically optimal in the sense of Huber (1961, 1972) for models which are similar to moving average models.

10.2. Nonparametric tests

Not many results are available as regards nonparametric tests for stochastic processes. The main work in this area is due to Bell, Woodroofe and Avadhani

(1970) where they obtain distribution-free tests for processes with 'independent structure', i.e., the processes which are either mixtures of i.i.d. processes or from which one can generate i.i.d. random variables.

Another aspect of the problem is to study the performance of the usual non-parametric tests under dependence. Serfling (1968) studied the Wilcoxon test when the dependence is of ϕ-mixing type. We shall describe these results briefly.

Suppose $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independent sample paths of a stationary ϕ-mixing process observed up to times 'm' and 'n' respectively with marginals F and G. The problem is to test the hypothesis $H: F = G$. The Wilcoxon statistic

$$U = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \varepsilon(Y_j - X_i),$$

where $\varepsilon(u) = -1, 0, 1$ according as $u < 0, = 0, > 0$ is used when the observations are i.i.d. One of the useful properties of the Wilcoxon two-sample statistic in the i.i.d. case is its asymptotic normality. Serfling (1968) proved that this property still holds even under dependence of ϕ-mixing type. It would be interesting to study the performance of other nonparametric tests under dependence. Albers (1977) has studied one-sample linear rank tests when the observations are from an autoregressive process. He showed that these tests applied to certain transformed observations have asymptotically the same properties as under independence both under the null hypothesis and under contiguous alternatives. Similar work has also been discussed in Modestino (1969).

11. Sequential inference

The main characteristic feature of sequential procedures is that the number of observations required or the time required for observation of the process is not determined in advance. The decision to terminate the observation on the process depends at each stage on the result of the observations made up to that stage. A merit of the sequential method is that test procedures and estimation procedures can be derived which require on the average a substantially smaller time period of observation or a smaller number of observations than equally reliable procedures based on a predetermined time of observation (or number of observations).

11.1. Sequential estimation

Consider a discrete time process as in Section 5. Let $\hat{\theta}_n$ be a MLE of θ. It can be shown that the method of maximum likelihood can be used with a suitable stopping rule to obtain estimators of θ with generalized variance, i.e., the determinant of the covariance matrix of $\hat{\theta}_n$ (when θ is a vector) smaller than any preassigned number. This makes use of the random central limit theorem for martingales. For details see

Basawa and Prakasa Rao (1980). One can obtain asymptotic confidence ellipsoids of given volume and given confidence coefficient by using sequential estimators of the above type. Results of this type were obtained by Anscombe (1952, 1953) for i.i.d. observations, and by Sarma (1976) for discrete time stationary Markov processes.

Sequential estimation problems for continuous time parameter stochastic processes with stationary independent increments were first studied by Dvoretsky, Kiefer and Wolfowitz (1953a, b).

Let $\{X(t, \theta); t \geq 0\}$, $\theta \in \Theta$ be a family of stochastic processes. Let $c(t)$ be the cost of observing the process up to time t and $W(\theta, \phi)$ the loss function on $\Theta \times \Theta$. Let T be a stopping time (cf. Doob (1953)) corresponding to the process and δ a decision rule depending on $X(t)$ only through its values for $0 \leq t \leq T$. The rule corresponding to T and δ is as follows: observe the process up to time T and then adopt the estimate δ. Such a procedure is called a *sequential plan*. One can study properties of minimax rules and Bayes rules among such sequential plans. Kiefer (1975) discussed invariant, minimax sequential plans when the invariance is with respect to either the decision problem or the time parameter.

In order to study optimality properties of a class of unbiased sequential plans, one can obtain an analogue of the Cramér–Rao inequality. An important lemma due to Sudakov (1969) is basic to the study of sequential problems for continuous parameter processes.

The significance of Sudakov's result is that it gives conditions under which stopped processes preserve absolute continuity and it provides a way of calculating their Radon–Nikodym derivative. Using this result, Magiera (1974) has derived an analogue of the Cramér–Rao inequality (see Rao (1973)) under some regularity conditions. For results of similar nature, see Trybula (1968), Franz and Winkler (1976) and Kagan, Linnik and Rao (1973). A short survey of the results is given in Winkler (1977).

11.2. Sequential tests

Suppose $\{Z(t), t \geq 0\}$ is a stationary process with independent increments with $Z(0) = 0$ defined on a probability space. Let $\mathscr{F}_t$ be the σ-algebra generated by $\{Z(s): 0 \leq s \leq t\}$. Suppose T is a stopping time with respect to the family $\{\mathscr{F}_t, t \geq 0\}$. Let

$$M(t) = \mathbf{E}[\mathrm{e}^{-tZ(T)}]$$

and $I = \{t: M(t) \geq 1\}$. If I is non-degenerate or if $\mathbf{E}\ T < \infty$ and $\mathbf{E}[Z(1)] = 0$, then it is known that (cf. Hall (1970))

$$\mathbf{E}[Z(T)] = \mathbf{E}(Z(1))\mathbf{E}[T].$$

This is an extension of Wald's equation for continuous time processes. Similar results of this nature, based on higher order moments, can be found in Hall (1969, 1970). The above relation can also be derived using the theory of martingales (cf. Doob

(1952), p. 380). Also, one can obtain a continuous time version of Wald's fundamental identity viz.

$$E[e^{-tZ(T)}M(t)^{-T}] = 1.$$

Shepp (1969) derived the above identity for Wiener processes.

Sequential probability ratio tests (SPRT) for continuous time processes with independent increments were first studied by Dvoretsky, Kiefer and Wolfowitz (1953a). Andrieu (1975) and Schmitz (1968, 1970) discussed SPRT tests for continuous and discrete time Markov processes. SPRT tests for discrete time stochastic processes were investigated by Eisenberg, Ghosh, Simmons (1976). Flavigny (1975) considered ϕ-mixing processes.

Sequential Bayes tests for the sign of the drift of Wiener process were studied by Chernoff (1961) in a series of papers. Bickel and Yahav (1972) are concerned with the problem of testing H_0: $\mu \leq 0$ versus H_1: $\mu > 0$ by sequential sampling from a one parameter exponential family of densities $f(x, \mu)$ with cost c per observation and zero one loss structure. These results are extensions of those of Chernoff (1961).

Sequential procedures for detecting parameter changes in time series models and for detection of increase in drift for Wiener process were discussed by Bagshaw and Johnson (1975a, b).

Acknowledgment

We wish to thank the editor and the referee for useful suggestions which greatly improved the presentation.

References

Albers, W. (1978), One sample rank tests under auto-regressive dependence, Ann. Statist. 6, 836–845.

Akritas, M. (1978), Contiguity of probability measures associated with continuous time stochastic processes, PhD thesis, University of Wisconsin, Madison.

Andrieu, C (1975), Une formulation, et la finitude, du test sequential du rapport des vraisemblaces dans le cas Markovien a temps continu, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 32, 333–344.

Anscombe, F.J. (1952), Large sample theory of sequential estimation, Proc. Cambridge Philos. Soc. 40, 600–607.

Anscombe, F.J. (1953), Sequential estimation, J. Roy. Statist. Soc. Ser. B 15, 1–29.

Athreya, K.B. and Keiding, N. (1977), Estimation theory for continuous time branching processes, Sankhyā Ser. A 39, 101–123.

Bagshaw, B. and Johnson, R.A. (1975a), Sequential detection of a drift change in Wiener process, Comm. Statist. 4, 787–796.

Bagshaw, B. and Johnson, R.A. (1975b), Sequential procedures for detecting parameter changes in a time series model, Technical Report, University Wisconsin, Madison.

Bahadur, R.R. (1964), On Fisher's bound for asymptotic variances, Ann. Math. Statist. 35, 1545–1552.

Banon, G. (1977), Estimation non parametrique de densite de probabilite pour les processus de Markov, Thesis, Université Paul Sabatier de Toulouse (France).

Banon, G. (1978), Nonparametric identification for diffusion processes, SIAM J. Control and Optimization 16, 380–395.

Bar-Shalom, Y. (1971), On the asymptotic properties of maximum likelihood estimate, obtained from dependent observations, J. Roy. Statist. Soc. Ser. B 33, 72–77.

Bartlett, M.S. (1955), Stochastic Processes, Cambridge University Press, London.

Basawa, I.V. (1974), Maximum likelihood estimation of parameters in renewal and Markov-renewal processes, J. Austr. Statist. Soc. 16, 33–43.

Basawa, I.V. (1977), Asymptotic statistical curvature for dependent data, Bull. Internat. Statist. Inst. 41st Session, Book 4, 45–48.

Basawa, I.V. (1980), Remarks on Bahadur efficiency of conditional tests, Ann. Statist., to appear.

Basawa, I.V. and Brockwell, P.J. (1978), Inference tor gamma and stable processes, Biometrika 65, 129–133.

Basawa, I.V. and Brockwell, P.J. (1980), A note on estimation for gamma and stable processes, Biometrika 67, 234–236.

Basawa, I.V., Feigin, P.D. and Heyde, C.C. (1976), Asymptotic properties of maximum likelihood estimators for stochastic processes, Sankhyā Ser. A 38, 259–270.

Basawa, I.V. and Koul, H.L. (1979), Asymptotic tests of composite hypotheses for non-ergodic type stochastic processes, Stochastic Processes Appl. 9, 291–305.

Basawa, I.V. and Koul, H.L. (1980), Asymptotically minimax tests of composite hypotheses for non-ergodic type stochastic processes, Preliminary Report, La Trobe University.

Basawa, I.V. and Prakasa Rao, B.L.S. (1980), Statistical Inference for Stochastic Processes, Academic, London, forthcoming.

Basawa, I.V. and Scott, D.J. (1976), Efficient tests for branching processes, Biometrika 63, 531–536.

Basawa, I.V. and Scott, D.J. (1977), Efficient tests for stochastic processes, Sankhyā Ser. A 39, 21–31.

Basawa, I.V. and Scott, D.J. (1978), Efficient estimation for stochastic processes, J. Indian Statist. Assoc., to appear.

Basawa, I.V. and Scott, D.J. (1979), Remarks on an asymptotic optimality criterion for estimation in stochastic processes, Sankhyā Ser. A, to appear.

Bechhofer, R.E., Kiefer, J. and Sobel, M. (1968), Sequential Identification and Ranking Procedures, Chicago University Press, Chicago.

Bell, C., Woodroofe, M. and Avadhani, T.V. (1970), Some non-parametric tests for stochastic processes, in: Non-Parametric Techniques in Statistical Inference, Cambridge University Press, London.

Bhat, B.R. (1974), On the method of maximum likelihood for dependent observations, J. Roy. Statist. Soc. Ser. B 36, 48–53

Bhat, B.R. and Kulkarni, S.R. (1972), Optimal asymptotic tests of composite hypotheses for stochastic processes, J. Karnatak Univ. Sci. 27, 73–89.

Bickel, P.J. and Doksum, K.A. (1977), Mathematical Statistics, Holden-Day, San Francisco.

Bickel, P.J. and Lehmann, E. (1969), Unbiased estimation in convex families, Ann. Math. Statist. 40, 1523–1535.

Bickel, P.J. and Yahav, J. (1969), Some contributions to the asymptotic theory of Bayes solutions, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 11, 257–275

Bickel, P. and Yahave, J. (1971), On the Wiener process approximation to Bayesian sequential testing problems, Proc. 6th Berkeley Symposium 1, 57–83.

Billingsley, P. (1961), Statistical Inference for Markov Processes, Chicago University Press, Chicago.

Borwanker, J.D. (1967), Some asymptotic results for stationary processes, PhD thesis, University of Minnesota.

Borwanker, J.D., Kallianpur, G., and Prakasa Rao, B.L.S. (1971), The Bernstein–Von Mises theorem for Markov processes, Ann. Math. Statist. 42, 1241–1253.

Brockwell, P.J. and Chung, K.L. (1975), Emptiness times of a dam with stable input and general release function, J. Appl. Probability 12, 212–217.

Brown, B.M. and Hewitt, J.I. (1975), Asymptotic likelihood theory for diffusion processes, J. Appl. Probability 12, 228–238.

Chernoff, H. (1961), Sequential tests for the mean of a normal distribution, Proc. 4th Berkeley Symposium 1, 79–91.

Cox, D.R. and Lewis, P.A.W. (1966), Statistical Analysis of Series of Events, Methuen, London.

Cramér, H. (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton.

Crane, M.A. and Lemoine, A.J. (1977), An Introduction to the Regenerative Method for Simulation Analysis, Springer, New York.
Crowder, M.J. (1976), Maximum likelihood estimation for dependent observations, J. Roy. Statist. Soc. Ser. B 38, 45–53.
Davies, R.B. (1978), Asymptotic inference when the amount of information is random, Technical Report, Wellington University.
Delecroix, M. (1975), Sur l'estimation des densites marginales et de transition d'un processus stationnaire et melangeant, Doctoral thesis, Université des Sciences et Techniques de Lille.
Doob, J.L. (1949), Applications of the theory of martingales, Colloq. Internat. C.N.R.S., Paris, 22–28.
Doob, J.L. (1953), Stochastic Processes, Wiley, New York.
Dorugovcev, A. Ya (1974), The consistency of an estimate of a parameter of a stochastic differential equation, Theory Probability and Math. Statist. 10, 73–82.
Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1953), Sequential decision problems with continuous time parameter: testing hypotheses, Ann. Math. Statist. 24, 254–264.
Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1953), Sequential decision Problems for processes with continuous time parameter: problems of estimation, Ann. Math. Statist. 24, 403–415.
Dzhaparidge, K.O. (1977), Tests of composite hypotheses for random variables and stochastic processes, Theor. Probability Appl. 22, 104–118.
Eisenberg, B., Ghosh, B.K. and Simons, G. (1976), Properties of generalized sequential probability ratio tests, Ann. Statist. 4, 237–251.
Feigin, P.D. (1975), Maximum likelihood estimation for stochastic processes – a martingale approach, PhD thesis, Australian National University.
Feigin, P.D. (1976), Maximum likelihood estimation for continuous time stochastic processes, Adv. Appl. Probability 8, 712–736.
Feigin, P.D. (1978), The efficiency criteria problem for stochastic processes, Stochastic Processes Appl. 6, 115–127.
Fisher, R.A. (1922), On the mathematical foundations of theoretical statistics, Philos. Trans. Roy. Soc. London Ser. A 222, 309–368.
Fisher, R.A. (1925), Theory of statistical estimation, Proc. Cambridge Philos. Soc. 22, 700–725.
Flavigny, R. (1975), A propos d'un test sequential concernant une suite ϕ-melangeante, C.R. Acad. Sci. Paris Ser. A 280, 1633–1636.
Foutz, R. (1974), Studies in large sample theory, PhD thesis, Ohio State University.
Franz, J. and Winkler, W. (1976), Über Stoppzeiten bei statistischen Problemen fur homogene Prozesse mit unabhangigen Zuwochsen, Math. Nachr. 70, 37–53.
Gastwirth, J.L., Rubin, H. and Wolff, S.S. (1961), The effect of autoregressive dependence on a non-parametric test, IEEE Trans. Information Theory 13, 311–313.
Gastwirth, J.L. and Rubin, H. (1971), Effect of dependence on the level of some one-sample tests, J. Amer. Statist. Assoc. 66, 816–820.
Gastwirth, J.L. and Rubin, H. (1975a), Asymptotic distribution theory of the empiric c.d.f. for mixing stochastic processes, Ann. Statist. 3, 809–824.
Gastwirth, J.L. and Rubin, H. (1975b), The behavior of robust estimators on dependent data, Ann. Statist. 3, 1070–1100.
Gikhman, I.I. and Skorokhod, A.V. (1972), Stochastic Differential Equations, Springer, Berlin.
Grenander, V. (1951), Stochastic processes and statistical inference, Ark. Mat. 1, 195–277.
Hajek, J. (1971), Local asymptotic minimax and admissibility in estimation, Proc. 6th Berkeley Symposium 1, 175–194.
Hall, W.J. (1969), Embedding submartingales in Wiener processes with drift, with applications to sequential analysis, J. Appl. Probability 6, 612–632.
Hall, W.J. (1970), On Wald's equations in continuous time, J. Appl. Probability 7, 59–68.
Heyde, C.C. (1975), Remarks on efficiency in estimation for branching processes, Biometrika 62, 49–55.
Heyde, C.C. and Feigin, P.D. (1975), On efficiency and exponential families in stochastic process estimation, in: E.C. Patel, S. Kolz and J. Ord, Eds., Statistical Distributions in Scientific Work, Vol. 1, 227–40, Reidel, Utrecht.
Heyde, C.C. (1978), On an optimal property of the maximum likelihood estimator of a parameter from a stochastic process, Stochastic Processes Appl. 1, 1–9.
Huber, P.J. (1964), Robust estimation of a location parameter, Ann. Math. Statist. 35, 73–101.

Huber, P.J. (1972), Robust statistics: a review, Ann. Math. Statist. 43, 1041–1067.

Ibragimov, I.A. and Khasminskii, R.Z. (1972), Asymptotic behavior of statistical estimators in the smooth case, I: study of the likelihood ratio, Theor. Probability Appl. 17, 445–462.

Ibragimov, I.A. and Khasminskii, R.Z. (1977), A problem of statistical estimation in Gaussian white noise, Soviet Math. Dokl. 18.

Jeganathan, P. (1979a), On a decomposition of the limit distribution of a sequence of estimators, Sankhyā Ser. A, to appear.

Jeganathan, P. (1979b), An extension of a result of L. LeCam concerning asymptotic normality, Res. Rep. Indian Statist. Inst.

Kagan, A.M., Linnik, Yu. V. and Rav, C.R. (1973), Characterization Problems in Mathematical Statistics, Wiley, New York.

Kailath, T.C. and Zakai, M. (1971), Absolute continuity and Radon–Nikodym derivatives for certain measures relative to Wiener measure, Ann. Math. Statist. 42, 130–140.

Keiding, N. (1974), Estimation in birth processes, Biometrika 61, 71–80.

Keiding, N. (1975), Maximum likelihood estimation in birth and death processes, Ann. Statist. 3, 363–372.

Kiefer, J. (1957), Some properties of generalized sequential probability ratio tests, Ann. Math. Statist. 28 573–601.

Koul, H.L. (1977), Behavior of robust estimators in the regression model with dependent errors, Ann. Statist. 5, 681–699.

LeCam, L. (1953), On some asymptotic properties of maximum likelihood and related Bayes estimates, Univ. Calif. Publ. in Statist. 1, 277–330.

Le Cam, L. (1960), *Locally asymptotically normal families of distributions*, Univ. Calif. Publ. Statist. 3, 27–98.

Le Cam, L. (1974), *Notes on asymptotic methods in statistical decision theory*, University Montreal Publication.

Lehmann, E.L. (1959), Testing Statistical Hypotheses, Wiley, New York.

Levit, B. Ya. (1974), On the behaviour of the generalised Bayes estimators for Markov observations. Theor. Probability Appl. 19, 332–346.

Magiera, R. (1974), On the inequality of Cramér-Rao type in sequential estimation theory, Zastos. Mat. Appl. Math. 14, 227–235.

M. Moore (1976), Remarks on the asymptotic behavior of posterior laws with applications, Utilitas Math. II, 237–259.

Modestino, J.W. (1969), Non-parametric and adaptive detecting on dependent data, Technical Report, Electrical Engineering Department, Princeton University.

Lipcer, R.S. and Sirjaev, A.N. (1972), On the absolute continuity of measures corresponding to processes of diffusion type, Math. USSR Izv. 6, 839–882.

Neyman, J. and Pearson, E.S. (1928), On the use and interpretation of certain test criteria for purposes of statistical reference, Biometrika 20A, 175–240, and 263–294.

Portnoy, S.D. (1977), Robust estimation in dependent situations, Ann. Statist. 5, 22–43.

Prakasa Rao, B.L.S. (1972), Maximum likelihood estimation for Markov processes, Ann. Inst. Statist. Math. 24, 333–345.

Prakasa Rao, B.L.S. (1973), On the rate of convergence of estimators for Markov processes, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 26, 141–152.

Prakasa Rao, B.L.S. (1974), Statistical Inference for Stochastic Processes, Technical Report, Université de Montreal.

Prakasa Rao, B.L.S. (1975), The Berry–Esseen bound for minimum contrast estimators in the independent not identically distributed case, Metrika 22, 225–239.

Prakasa Rao, B.L.S. (1977), Berry–Esseen type bound for density estimators of stationary Markov processes, Bull. Math. Statist. 17, 15–21.

Prakasa Rao, B.L.S. (1978), Density estimation for Markov processes using Delta-sequences, Ann. Inst. Statist. Math. 30, 321–328.

Prakasa Rao, B.L.S. (1978), On some problems of inference for Markov processes, Proc. Gen. Math. Seminar, University of Patras, Greece, to appear.

Prakasa Rao, B.L.S. (1978), The rate of convergence of Bernstein–Von Mises approximation for Markov processes, Serdica 4, 36–42.

Prakasa Rao, B.L.S. (1979), Nonparametric density estimation for continuous time Markov processes via delta families, Publ. Inst. Statist. Univ. Paris 24, 79–97.

Prakasa Rao, B.L.S. (1979), Asymptotic theory for non-linear least squares estimators for diffusion processes, Technical Report, University of Poona.

Prakasa Rao, B.L.S. (1979), The equivalence of Bayes and maximum likelihood estimators for Markov processes, Ann. Inst. Statist. Math., 31, 179–193.

Prakasa Rao, B.L.S. (1980), Maximum probability estimation for diffusion processes, preprint, University of Wisconsin, Madison.

Prakasa Rao, B.L.S. and Rubin, H. (1979), Asymptotic theory for process least squares estimators for diffusion processes, Technical Report, Purdue University.

Prasad, M.S. (1973), Some contribution to the theory of maximum likelihood estimation for dependent random variables, PhD thesis, Indian Institute of Technology, Kanpur.

Prasad, M.S. and Prakasa Rao, B.L.S. (1976), Maximum likelihood estimation for dependent random variables, J. Indian Statist. Assoc. 14, 75–97.

Ranneby, B. (1975), Robustness against dependence of maximum likelihood estimates in i.i.d. models, Technical Report, University of Umea, Sweden.

Ranneby, B. (1975), Maximum likelihood estimation for continuous time Markov chains, Technical Report, University of Umea, Sweden.

Rao, C.R. (1961), Asymptotic efficiency and limiting information, Proc. 4th Berkeley Symposium **1**, 531–545.

Rao, C.R. (1962), Efficient estimates and optimum inference procedures in large samples, J. Roy. Statist. Ser. B 24, 46–72.

Rao, C.R. (1963), Criteria of estimation in large samples, Sankhyā 25, 189–206.

Rao, C.R. (1973), Linear Statistical Inference (2nd Edition), Wiley, New York.

Rao, M.M. (1963), Inference in stochastic processes I, Theor. Probability Appl. 8, 266–281.

Rao, M.M. (1966), Inference in stochastic processes II, Z. Wahracheinlichkeitstheorie und Verw. Gebiete 5, 317–335.

Roussas, G.G. (1968), On some properties of non-parametric estimates of probability density functions, Bull. Soc. Math. Grèce 2, 29–43.

Roussas, G.G. (1969a), Non-parametric estimation in Markov processes, Ann. Inst. Statist. Math. 21, 73–87.

Roussas, G.G. (1969b), Non-parametric estimation of the transition distribution function of a Markov process, Ann. Math. Statist. 40, 1386–1400.

Roussas, G.G. (1972), *Contiguity of Probability Measures*, Cambridge University Press, London.

Roussas, G.G. (1977), Asymptotic properties of the maximum probability estimates in Markov processes, Ann. Inst. Statist. Math. 29, 203–219.

Rosenblatt, M. (1956), Remarks on some non-parametric estimates of a density function, Ann. Math. Statist. 27, 832–837.

Rosenblatt, M. (1970), Density estimates and Markov sequences, in: Non-Parametric Techniques in Statistical Inference, Cambridge University Press, London.

Sarma, Y.R. (1976), Large sample theory of sequential estimation in stationary Markov processes, Pub. Inst. Statist. Univ. Paris 21, 57–70.

Schmitz, N. (1968), Likelihood quotienten—sequenztests bei homogenen Markovschen Ketten, Biometriche Z., IQ, 231–347.

Schmitz, N. (1970), Existenz von sequentiellen Tests zu vorgegbenen Niveaus, Arch. Math. 21, 617–628.

Schwartz, L. (1965), On Bayes procedures, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 4, 10–26.

Sen, P.K. (1977), On Wiener process embedding for linear combination of order statistics, Sankhyā Ser. A 39, 138–143.

Sen, P.K. and Ghush, M. (1973), A Chernoff–Savage representation of rank order statistics for stationary ϕ-mixing processes, Sankhyā Ser. A 35, 153–172.

Serfling, R.J. (1968), The Wilcoxon two-sample statistic on strongly mixing processes, Ann. Math. Statist. 39, 1202–1209.

Shepp, L. (1969), Explicit solutions to some problems of optimal stopping, Ann. Math. Statist. 40, 993–1010.

Silvey, S.D. (1961), A note on the maximum likelihood in the case of dependent random variables, J. Roy. Statist. Soc. Ser. B, 23 444–452.

Strasser, H. (1977), Improved bounds for the equivalence of Bayes and maximum likelihood estimation, Teor. Verojathost. Primenen 22, 358–370.

Strasser, H. (1977), Asymptotic properties of posterior distributions, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 35, 269–282.

Sudakov, V.N. (1969), On measures defined by Markovian moments, in: Investigations on the Theory of Random Processes, Vol. 12, Memoirs of Scientific Seminars of the Leningrad Section of Steklov Math. Inst., 157–164.

Sweeting, T.J. (1978), On efficient tests for branching processes, Biometrika 65, 123–127.

Taraskin, A. (1974), On the asymptotic normality of a vector-valued stochastic integral and estimates of drift parameters of a multi-dimensional diffusion process, Theory Probability and Math. Statist. 2, 209–224.

Thomirov, V.M. (1960), Russian Math. Surveys 15.

Trybula, S. (1968), Sequential estimation in processes with independent increments, Dissertationes Math. 50.

Wald, A. (1939), Contributions to the theory of statistical estimation and testing hypotheses, Ann. Math. Stat. 10, 299–326.

Wald, A. (1948), Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process, Ann. Math. Statist. 20, 595–601.

Wald, A. (1949), Note on the consistency of the maximum likelihood estimate, Ann. Math. Statist. 20, [illegible]–601.

Wald, A. (1950), Statistical Decision Functions, Wiley, New York.

Wald, (1958), Selected Papers in Statistics and Probability by Abraham Wald, Stanford University Press, Stanford.

Weiss, L. (1971), Asymptotic properties of maximum likelihood estimators in some non-standard cases, J. Amer. Statist. Assoc. 66, 345–350.

Weiss, L (1973), Asymptotic properties of maximum likelihood estimates in some non-standard cases II, J. Amer. Statist. Assoc. 68, 428–430.

Weiss, L (1975), The asymptotic distribution of the likelihood ratio in some nonstandard cases, J. Amer. Statist. Assoc. 70, 204–208.

Weiss, L. and Wolfowitz, J. (1966), Generalized maximum likelihood estimators, Theor. Probability Appl. 11, 58–81.

Weiss, L. and Wolfowitz, J. (1969), Asymptotically minimax tests of composite hypotheses, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 14, 161–168.

Weiss, L. and Wolfowitz, J. (1974), Maximum Probability Estimators and Related Topics, Lecture Notes in Mathematics 424, Springer, Berlin.

Wilks, S.S. (1944), *Mathematical Statistics*, Princeton University Press, Princeton.

Winkler, W. (1977), Sequential estimation in processes with independent increments, Banach Center Pub 6, to appear.

Wolff, R.W. (1965), Problems of statistical inference for birth and death queueing models, Operations Res. 13, 343–357.

Yamada, K. (1976), Asymptotic behavior of posterior distributions for random processes under incorrect models, J. Math. Anal. Appl. 56, 294–308.